GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 10:04:34 ; Search time 23252 Seconds
(without alignments)
525.683 Million cell updates/sec

Title: US-09-696-664A-3
Perfect score: 420
Sequence: 1 gcgcgttgtgagaattctta.....tcattgtgggtaacgtattt 420

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

[redacted]rched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		408.6	97.3	140384	8	ZMA86563	X86563 Zea mays co
2		399.4	95.1	1990	8	CHNMRBCL	X55828 Neurachne m
3		399.4	95.1	2010	8	CHNTRBCL	X55827 Neurachne t
4		392.2	93.4	1990	8	SILCRBCL	X79900 S.italica L
5	c	389.8	92.8	183990	8	AF114171	AF114171 Sorghum b
6		386	91.9	1803	8	CHZM02	V00171 Zea mays ch
7		385	91.7	1989	8	CHOSRBCL	X04789 Rice chloro
8		385	91.7	2062	8	RICCPRBCL	D00207 Oryza sativ
9		385	91.7	125936	8	AC122148	AC122148 Oryza sat
10		385	91.7	134525	8	CHOSXX	X15901 Oryza sativ
11		385	91.7	134933	8	AC092750	AC092750 Oryza sat
12		385	91.7	147201	8	AP003280	AP003280 Oryza sat
13		383.4	91.3	105649	2	AP003986	AP003986 Oryza sat
14	c	383.4	91.3	139914	2	AP003623	AP003623 Oryza sat
15	c	383.4	91.3	167379	2	AP004236	AP004236 Oryza sat
16		380.2	90.5	139671	2	CNS07EFJ	AL513004 Oryza sat
17		377	89.8	1355	8	RICCHRBCLA	L24073 Oryza sativ
18		376.6	89.7	749	8	AY090549	AY090549 Deschamps
19		376.6	89.7	1623	8	ASTCPRBCL	L15300 Avena sativ
20		370.2	88.1	1431	8	CHZMRBCLG	Z11973 Z.mays chlo
21		368.6	87.8	1755	8	WHTCPRBCL	D00206 Triticum ae
22		368.6	87.8	2174	8	CHHV02	X00630 Barley chlo
23		368.6	87.8	134545	8	AB042240	AB042240 Triticum
24		357.4	85.1	1510	8	CEHRBCL	L14622 Cenchrus se
25		354.2	84.3	1430	8	PENCARBOXL	L14623 Pennisetum
26		354.2	84.3	2145	8	CHCLARBCL	X69731 C.laevigatu
27		348.8	83.0	2126	8	CHPCORBCL	X69748 P.comosa ch
28		345.6	82.3	1839	8	CHFBRBCL	X55830 Flaveria bi
29		345.6	82.3	1842	8	CHFPRBCL	X55829 Flaveria pr
30		344	81.9	2088	8	MTAPRBCL	X83986 A.punctatum
31		344	81.9	2283	8	CHGHRBCL	X15886 Cotton chlo
32		342.4	81.5	2087	8	MTSARBCL	X83996 S.australe
33	c	342.4	81.5	163935	8	OEL271079	AJ271079 Oenothera
34		341.4	81.3	2193	8	IPRBCL	X60663 I.purpurea
35		340.8	81.1	1661	8	IEXCPRBCL	M88583 Ilex vomito
36		340.8	81.1	2052	8	TOBCPRUBA	M16867 Tobacco (N.
37		340.8	81.1	2052	8	TOBCPRUBB	M16896 Tobacco (N.
38		340.8	81.1	2524	8	TOBCPRUBP	J01450 Nicotiana t
39		340.8	81.1	4260	6	AX137516	AX137516 Sequence
40		340.8	81.1	4260	6	E51181	E51181 Method for
41		340.8	81.1	4260	6	E51199	E51199 Process for
42		340.8	81.1	155939	8	CHNTXX	Z00044 Nicotiana t
43		340.2	81.0	1344	8	TBU31381	U31381 Tristachya
44		339.8	80.9	2074	8	CHNMURBCL	X69747 N.mucronatu
45		339.2	80.8	2172	8	CHHRRBCL	X69740 L.africana

ALIGNMENTS

RESULT 1
ZMA86563
LOCUS ZMA86563 140384 bp DNA circular PLN 12-APR-2001
DEFINITION Zea mays complete chloroplast genome.
ACCESSION X86563
VERSION X86563.2 GI:11990232
KEYWORDS 16S rRNA; 23S rRNA; 4.5S rRNA; 5S rRNA; ATPase; atpE gene; atpI gene; cemA gene; clpP gene; cytochrome b/f complex; cytochrome b6; cytochrome f; infA gene; initiation factor 1; junction LSC-IR; matK gene; NADH dehydrogenase; ndhB gene; ndhC gene; ndhD gene; ndhF gene; ndhJ gene; ndhK gene; petA gene; petB gene; petD gene; petE gene; psaA gene; psaB gene; psaI gene; psaJ gene; psbA gene; psbB gene; psbC gene; psbD gene; psbF gene; psbH gene; psbI gene; psbJ gene; psbK gene; psbN gene; PSI P700 apoprotein A1; PSI P700 apoprotein A2; PSII component; PSII cytochrome b559; PSII D2

protein; PSII I protein; PSII K protein; PSII L protein; PSII protein; rbcL gene; ribosomal protein L14; ribosomal protein L16; ribosomal protein L2; ribosomal protein L20; ribosomal protein L22; ribosomal protein L23; ribosomal protein L32; ribosomal protein L33; ribosomal protein L36; ribosomal protein S11; ribosomal protein S12; ribosomal protein S14; ribosomal protein S15; ribosomal protein S16; ribosomal protein S18; ribosomal protein S19; ribosomal protein S2; ribosomal protein S3; ribosomal protein S4; ribosomal protein S7; ribosomal protein S8; ribosomal RNA; RNA polymerase; RNA-Asparagine; rpl14 gene; rpl16 gene; rpl2 gene; rpl20 gene; rpl22 gene; rpl23 gene; rpl32 gene; rpl33 gene; rpl36 gene; rpoA gene; rpoB gene; rpoC1 gene; rpoC2 gene; rps11 gene; rps12 gene; rps14 gene; rps15 gene; rps16; rps18 gene; rps19 gene; rps2 gene; rps3 gene; RPS4 gene; rps7 gene; rps8 gene; Rubisco; transfer RNA-Ala; transfer RNA-Arg; transfer RNA-Asp; transfer RNA-Cys; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His; transfer RNA-Iso; transfer RNA-Leu; transfer RNA-Lys; transfer RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser; transfer RNA-Thr; transfer RNA-Try; transfer RNA-Tyr; transfer RNA-Val.

SOURCE Zea mays.
ORGANISM Plastid Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 140384)
AUTHORS Strittmatter,G. and Kossel,H.
TITLE Cotranscription and processing of 23S, 4.5S and 5S rRNA in chloroplasts from Zea mays
JOURNAL Nucleic Acids Res. 12 (20), 7633-7647 (1984)
MEDLINE 85037934
PUBMED 6093045

REFERENCE 2 (bases 1 to 140384)
AUTHORS Erion,J.L.
TITLE Characterization of the mRNA transcripts of the maize, ribulose-1,5-bisphosphate carboxylase, large subunit gene
JOURNAL Plant Mol. Biol. 4, 169-179 (1985)

REFERENCE 3 (bases 1 to 140384)
AUTHORS Fish,L.E., Kuck,U. and Bogorad,L.
TITLE Two partially homologous adjacent light-inducible maize chloroplast genes encoding polypeptides of the P700 chlorophyll a-protein complex of photosystem I
JOURNAL J. Biol. Chem. 260 (3), 1413-1421 (1985)
MEDLINE 85104999
PUBMED 3881431

REFERENCE 4 (bases 1 to 140384)
AUTHORS Rock,C.D., Barkan,A. and Taylor,W.C.
TITLE The maize plastid psbB-psbF-petB-petD gene cluster: spliced and unspliced petB and petD RNAs encode alternative products
JOURNAL Curr. Genet. 12 (1), 69-77 (1987)
MEDLINE 88210525
PUBMED 2835175

REFERENCE 5 (bases 1 to 140384)
AUTHORS McLaughlin,W.E. and Larrinua,I.M.
TITLE The sequence of the maize plastid encoded rpl 22 locus
JOURNAL Nucleic Acids Res. 15 (10), 4356 (1987)
MEDLINE 87231016
PUBMED 3295778

REFERENCE 6 (bases 1 to 140384)
AUTHORS Rodermel,S., Orlin,P. and Bogorad,L.
TITLE The transcription termination region between two convergently-transcribed photoregulated operons in the maize plastid chromosome contains rps14, trnR (UCU) and a putative trnfM pseudogene
JOURNAL Nucleic Acids Res. 15 (13), 5493 (1987)
MEDLINE 87260027
PUBMED 3601681

REFERENCE 7 (bases 1 to 140384)
AUTHORS Giese,K., Subramanian,A.R., Larrinua,I.M. and Bogorad,L.
TITLE Nucleotide sequence, promoter analysis, and linkage mapping of the unusually organized operon encoding ribosomal proteins S7 and S12 in maize chloroplast
JOURNAL J. Biol. Chem. 262 (31), 15251-15255 (1987)
MEDLINE 88033114
PUBMED 2822717

REFERENCE 8 (bases 1 to 140384)
AUTHORS Lukens,J.H. and Bogorad,L.
TITLE Nucleotide sequence containing the maize chloroplast proline (UGG) and tryptophan (CCA) tRNA genes
JOURNAL Nucleic Acids Res. 16 (11), 5192 (1988)
MEDLINE 88262508
PUBMED 3387224

REFERENCE 9 (bases 1 to 140384)
AUTHORS Ruf,M. and Kossel,H.
TITLE Structure and expression of the gene coding for the alpha-subunit of DNA-dependent RNA polymerase from the chloroplast genome of Zea mays
JOURNAL Nucleic Acids Res. 16 (13), 5741-5754 (1988)
MEDLINE 88289331
PUBMED 3399379

REFERENCE 10 (bases 1 to 140384)
AUTHORS Bowman,C.M., Barker,R.F. and Dyer,T.A.
TITLE In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination
JOURNAL Curr. Genet. 14 (2), 127-136 (1988)
MEDLINE 89028843
PUBMED 3180271

REFERENCE 11 (bases 1 to 140384)
AUTHORS Schantz,R. and Bogorad,L.
TITLE Maize chloroplast genes ndhD, ndhE, and psaC. Sequences, transcripts and transcript pools
JOURNAL Plant Mol. Biol. 11, 239-247 (1988)

REFERENCE 12 (bases 1 to 140384)
AUTHORS Steinmuller,K., Ley,A.C., Steinmetz,A.A., Sayre,R.T. and Bogorad,L.
TITLE Characterization of the ndhC-psbG-ORF157/159 operon of maize plastid DNA and of the cyanobacterium Synechocystis sp. PCC6803
JOURNAL Mol. Gen. Genet. 216 (1), 60-69 (1989)
MEDLINE 89281491
PUBMED 2499764

REFERENCE 13 (bases 1 to 140384)
AUTHORS Igloi,G.L., Meinke,A., Dory,I. and Kossel,H.
TITLE Nucleotide sequence of the maize chloroplast rpo B/C1/C2 operon: comparison between the derived protein primary structures from various organisms with respect to functional domains
JOURNAL Mol. Gen. Genet. 221 (3), 379-394 (1990)
MEDLINE 90340289
PUBMED 2381419

REFERENCE 14 (bases 1 to 140384)
AUTHORS Fitzky,B. and Subramanian,A.R.
TITLE Nucleotide sequence and map positions of the duplicated gene for chloroplast ribosomal protein S15 in Zea mays (maize)
JOURNAL Nucleic Acids Res. 18 (11), 3407 (1990)
MEDLINE 90287730
PUBMED 2129550

REFERENCE 15 (bases 1 to 140384)
AUTHORS Kavousi,M., Giese,K., Larrinua,I.M., McLaughlin,W.E. and Subramanian,A.R.
TITLE Nucleotide sequence and map positions of the duplicated gene for maize (Zea mays) chloroplast ribosomal protein L2
JOURNAL Nucleic Acids Res. 18 (14), 4244 (1990)
MEDLINE 90332419
PUBMED 2377464

REFERENCE 16 (bases 1 to 140384)
AUTHORS Kangasjarvi,J., McCullough,A. and Gengenbach,B.G.
TITLE Nucleotide sequence and transcription of maize plastid genome Bam HI fragment 14 containing ORF170
JOURNAL Plant Mol. Biol. 17 (3), 513-515 (1991)
MEDLINE 91355944
PUBMED 1884003

REFERENCE 17 (bases 1 to 140384)
AUTHORS Weglohner,W. and Subramanian,A.R.
TITLE Nucleotide sequence of a region of maize chloroplast DNA containing the 3' end of clpP, exon 1 of rps12 and rpl20 and their cotranscription
JOURNAL Plant Mol. Biol. 18 (2), 415-418 (1992)
MEDLINE 92119264

```
  PUBMED   1732000
REFERENCE  18 (bases 1 to 140384)
  AUTHORS  Rodermel,S.R.
  TITLE    Nucleotide sequence of a maize chloroplast DNA fragment containing
           an inversion breakpoint, trnG (GCC), trnG (UCC), trnfM, and a trnG
           pseudogene
  JOURNAL  Nucleic Acids Res. 20 (21), 5844 (1992)
  MEDLINE  93087206
  PUBMED   1454547
REFERENCE  19 (bases 1 to 140384)
  AUTHORS  Maier,R.M., Neckermann,K., Hoch,B., Akhmedov,N.B. and Kossel,H.
  TITLE    Identification of editing positions in the ndhB transcript from
           maize chloroplasts reveals sequence similarities between editing
           sites of chloroplasts and plant mitochondria
  JOURNAL  Nucleic Acids Res. 20 (23), 6189-6194 (1992)
  MEDLINE  93117088
  PUBMED   1282235
REFERENCE  20 (bases 1 to 140384)
  AUTHORS  Weglohner,W. and Subramanian,A.R.
  TITLE    Nucleotide sequence of maize chloroplast rpl32: completing the
           apparent set of plastid ribosomal protein genes and their tentative
           operon organization
  JOURNAL  Plant Mol. Biol. 21 (3), 543-548 (1993)
  MEDLINE  93184210
  PUBMED   8443346
REFERENCE  21 (bases 1 to 140384)
  AUTHORS  Maier,R.M., Neckermann,K., Igloi,G.L. and Kossel,H.
  TITLE    Complete sequence of the maize chloroplast genome: gene content,
           hotspots of divergence and fine tuning of genetic information by
           transcript editing
  JOURNAL  J. Mol. Biol. 251 (5), 614-628 (1995)
  MEDLINE  95395841
  PUBMED   7666415
REFERENCE  22 (bases 1 to 140384)
  AUTHORS  Neckermann,K.
  TITLE    Direct Submission
  JOURNAL  Submitted (25-APR-1995) K. Neckermann, Institut f. Biologie III,
           Schaenzlestr. 1, D- 79104 Freiburg i.Br., FRG
  REMARK   revised by [22]
REFERENCE  23 (bases 1 to 140384)
  AUTHORS  Neckermann,K.

  Query Match             97.3%;  Score 408.6;  DB 8;  Length 140384;
  Best Local Similarity   99.0%;  Pred. No. 1.4e-104;
  Matches  411;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        6 TTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAAC 65
            ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    56834 TTGTGTGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAAC 56893

Qy       66 TAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCC 125
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    56894 TAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCC 56953

Qy      126 GGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGG 185
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    56954 GGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGG 57013

Qy      186 GGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGAC 245
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    57014 GGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGAC 57073

Qy      246 AACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACAT 305
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    57074 AACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACAT 57133

Qy      306 CGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCT 365
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    57134 CGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCT 57193

Qy      366 ATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
            |||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db    57194 ATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 57248
```

```
RESULT 2
CHNMRBCL
LOCUS       CHNMRBCL                1990 bp    DNA     linear   PLN 29-MAR-2001
DEFINITION  Neurachne munroi chloroplast rbcL gene for the rubisco large
            subunit (EC 4.1.1.39).
ACCESSION   X55828 M33797 X15921
VERSION     X55828.1  GI:11750
KEYWORDS    rbcL gene; ribulose bisphosphate carboxylase; rubisco large
            subunit.
SOURCE      Neurachne munroi.
  ORGANISM  Plastid Neurachne munroi
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Paniceae; Neurachne.
REFERENCE   1  (bases 1 to 1990)
  AUTHORS   Hudson,G.S., Mahon,J.D., Anderson,P.A., Gibbs,M.J., Badger,M.R.,
            Andrews,T.J. and Whitfeld,P.R.
  TITLE     Comparisons of rbcL genes for the large subunit of
            ribulose-bisphosphate carboxylase from closely related C3 and C4
            plant species
  JOURNAL   J. Biol. Chem. 265 (2), 808-814 (1990)
  MEDLINE   90110139
  PUBMED    2295620
REFERENCE   2  (bases 1 to 1990)
  AUTHORS   Hudson,G.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1989) Hudson G.S., Research School of Biological
            Sciences, PO Box 475, Australian National University, Canberra City
            ACT 2601, Australia
FEATURES             Location/Qualifiers
     source          1. .1990
                     /organism="Neurachne munroi"
                     /organelle="plastid"
                     /db_xref="taxon:4524"
     promoter        87. .92
                     /note="-35 box"
     promoter        111. .116
                     /note="-10 box"
     precursor_RNA   121. .>121
                     /note="put. rbcL transcript"
     CDS             418. .1854
                     /note="rubisco large subunit"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA39354.1"
                     /db_xref="GI:11751"
                     /db_xref="SWISS-PROT:P19163"
                     /translation="MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVTP
                     QPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPGEEDQYICYVA
                     YPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPTYSKTFQGPPHGIQVERD
                     KLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDENVNSQPFMRWRDRF
                     VFCAEAIYKAQAETGEIKGHYLNATAGTCEEMMKRAAFARELGVPIVMHDYLTGGFTA
                     NTSLSMYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHVHSGTVVGK
                     LEGEREITLGFVDLLRDDFIEKDRSRGVFFTQDWVSMPGVIPVASGGIHVWHMPALTE
                     IFGDDSVLQFGGGTLGHPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIKAACK
                     WSPELAAACEVWKAIKFEFEPVDTVDKV"
     repeat_unit     1941. .1982
                     /rpt_type=INVERTED
BASE COUNT      613 a    340 c    445 g    592 t
ORIGIN

  Query Match             95.1%;  Score 399.4;  DB 8;  Length 1990;
  Best Local Similarity   97.4%;  Pred. No. 4.6e-102;
  Matches  406;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy        4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
            |||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db      376 CGTTGTGAGAATTATTAATTAATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 435

Qy       64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    436 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 495

Qy    124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |
Db    496 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTTCGAGTAACTCCTCAGCCC 555

Qy    184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    556 GGGGTTCCGCCTGAAGAAGCAGGGGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 615

Qy    244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    616 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 675

Qy    304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    676 ATCGAGCCCGTTCCTGGGGAAGAAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 735

Qy    364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db    736 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 792
```

RESULT 3
CHNTRBCL

```
LOCUS       CHNTRBCL                2010 bp    DNA     linear   PLN 29-MAR-2001
DEFINITION  Neurachne tenuifolia chloroplast rbcL gene for the rubisco large
            subunit (EC 4.1.1.39).
ACCESSION   X55827 M33796 X15920
VERSION     X55827.1  GI:11797
KEYWORDS    rbcL gene; ribulose bisphosphate carboxylase; rubisco large
            subunit.
SOURCE      Neurachne tenuifolia.
  ORGANISM  Plastid Neurachne tenuifolia
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Paniceae; Neurachne.
REFERENCE   1  (bases 1 to 2010)
  AUTHORS   Hudson,G.S., Mahon,J.D., Anderson,P.A., Gibbs,M.J., Badger,M.R.,
            Andrews,T.J. and Whitfeld,P.R.
  TITLE     Comparisons of rbcL genes for the large subunit of
            ribulose-bisphosphate carboxylase from closely related C3 and C4
            plant species
  JOURNAL   J. Biol. Chem. 265 (2), 808-814 (1990)
  MEDLINE   90110139
   PUBMED   2295620
REFERENCE   2  (bases 1 to 2010)
  AUTHORS   Hudson,G.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1989) Hudson G.S., Research School of Biological
            Sciences, PO Box 475, Australian National University, Canberra City
            ACT 2601, Australia
FEATURES             Location/Qualifiers
     source          1. .2010
                     /organism="Neurachne tenuifolia"
                     /organelle="plastid"
                     /db_xref="taxon:4525"
     promoter        87. .92
                     /note="-35 box"
     promoter        111. .116
                     /note="-10 box"
     precursor_RNA   121. .>121
                     /note="put. rbcL transcript"
     CDS             438. .1874
                     /note="rubisco large subunit"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA39353.1"
                     /db_xref="GI:11798"
                     /db_xref="SWISS-PROT:P19164"
                     /translation="MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVTP
                     QPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPGEEDQYICYVA
                     YPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPTYSKTFQGPPHGIQVERD
                     KLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDENVNSQPFMRWRDRF
                     VFCAEAIYKAQAETGEIKGHYLNATAGTCEEMMKRAIFARELGVPIVMHDYLTGGFTA
                     NTSLAMYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHVHAGTVVGK
                     LEGEREITLGFVDLLRDDFIEKDRSRGVFFTQDWVSMPGVIPVASGGIHVWHMPALTE
                     IFGDDSVLQFGGGTLGHPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIKAACK
                     WSPELAAACEVWKAIKFEFAPVDTVDKV"
     repeat_unit     1961. .2002
                     /rpt_type=INVERTED
BASE COUNT      614 a    345 c    453 g    598 t
ORIGIN
```

```
  Query Match             95.1%;  Score 399.4;  DB 8;  Length 2010;
  Best Local Similarity   97.4%;  Pred. No. 4.6e-102;
  Matches  406;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          |||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||
Db    396 CGTTGTGAGAATTATTAATTAATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 455

Qy     64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 515

Qy    124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |
Db    516 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTTCGAGTAACTCCTCAGCCC 575

Qy    184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    576 GGGGTTCCGCCTGAAGAAGCAGGGGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 635

Qy    244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    636 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 695

Qy    304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    696 ATCGAGCCCGTTCCTGGGGAGGAAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 755

Qy    364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db    756 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 812
```

RESULT 4
SILCRBCL

```
LOCUS       SILCRBCL                1990 bp    DNA     linear   PLN 29-OCT-1996
DEFINITION  S.italica L. chloroplast rbcL gene.
ACCESSION   X79900
VERSION     X79900.1  GI:1654091
KEYWORDS    large subunit; rbcL gene; ribulose 1,5-bisphosphate.
SOURCE      Setaria italica.
  ORGANISM  Chloroplast Setaria italica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Paniceae; Setaria.
REFERENCE   1  (bases 1 to 1990)
  AUTHORS   Zhao,Y.S., Qiao,X.Y., Wu,N.H. and Wu,X.Y.
  TITLE     Nucleotide sequence of ribulose-1,5-bisphosphate
            carboxylase/oxygenase large subunit gene from millet (Setaria
            italica)
  JOURNAL   Acta Bot. Sin. 38, 719-724 (1996)
REFERENCE   2  (bases 1 to 1990)
  AUTHORS   Zhao,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1994) Y. Zhao, Institute of Developmental
            Biology, Academia Sinica, PO Box 2707, Beijing 100080, PROC
FEATURES             Location/Qualifiers
     source          1. .1990
                     /organism="Setaria italica"
                     /organelle="plastid:chloroplast"
                     /db_xref="taxon:4555"
                     /clone="pSIL"
```

```
     gene            105. .1990
                     /gene="rbcL"
     mRNA            105. .1990
                     /gene="rbcL"
                     /product="large subunit of ribulose 1,5-bisphosphate"
     RBS             380. .384
                     /gene="rbcL"
     mRNA            390. .1820
                     /gene="rbcL"
                     /product="large subunit of ribulose 1,5-bisphosphate"
BASE COUNT      579 a    359 c    451 g    601 t
ORIGIN

  Query Match             93.4%;  Score 392.2;  DB 8;  Length 1990;
  Best Local Similarity   97.8%;  Pred. No. 5e-100;
  Matches  408;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
         |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   349 CGTTGTGAGAATTCTTAATTCATGA-TTGTAGGGAGGGACTTATGTCACCACAAACAGAA 407

Qy    64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   408 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 467

Qy   124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   468 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCC 527

Qy   184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   528 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 587

Qy   244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   588 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 647

Qy   304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
         |||||||||||||||||||  |||||||||||||||||||| ||||||||||||||||||
Db   648 ATCGAGCCCGTTCCTGGGGAGGCAGATCAATATATCTGTTATATAGCTTATCCATTAGAC 707

Qy   364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
         ||||||||||||||||||  |||||||||||||||||| |||||||||||||||| |||
Db   708 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTGTTT 764


LT 5
14171/c
LOCUS       AF114171              183990 bp    DNA     linear   PLN 25-APR-1999
DEFINITION  Sorghum bicolor BAC clone 25.M18, complete sequence.
ACCESSION   AF114171
VERSION     AF114171.1  GI:4680196
KEYWORDS    .
SOURCE      Sorghum bicolor.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 183990)
  AUTHORS   Llaca,V., Lou,A., Young,S. and Messing,J.
  TITLE     Retrotransposable elements of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183990)
  AUTHORS   Llaca,V., Lou,A., Young,S. and Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1998) Waksman Institute, Rutgers University, 190
            Frelinghuysen Rd., Piscataway, NJ 08854-8020, USA
FEATURES             Location/Qualifiers
     source          1. .183990
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /chromosome="F"
                     /clone="BAC 25.M18"
                     /clone_lib="Woo et. al., Nucl. Acids Res. 22:4922-4931,"
     repeat_region   3517. .3644
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     mRNA            join(<3841. .4596,4699. .>5541)
                     /product="putative beta-ketoacyl-CoA synthase"
     CDS             join(3841. .4596,4699. .5541)
                     /note="similar to Oryza sativa EST clones E10604_1A,
                     E31307_5A, S3416_3A"
                     /codon_start=1
                     /product="putative beta-ketoacyl-CoA synthase"
                     /protein_id="AAD27560.1"
                     /db_xref="GI:4680197"
                     /translation="MSIKSETNERSSTIQSAGHRICSIVKLCYQHLMLAVSYALYLFL
                     APIILATLVLHLSRHTSADDDLIVSARSTLTANLPFTVVVVAAAATAVYLMRRPRAVY
                     LLDLACYKPGAEHAVTRETFMRLVVGYGTFSDDSLAFQRKILERSGLGQGTYLPKALT
                     SSPPNPCMAEARREAEAVMFGAVDQLLAKTGVRAGDIGIVVVNSGLFSPTPSLSAMIV
                     HHYKLRRDVSSYNLAGMGCSASLISVDLAKQLLQVQRNTYALVVSLEIMAASAYVGNT
                     RSMLVPNCIFRMGGAAVLLTNHGTERRRAKYQLLHTVRTHQGADDRAYRCVYQEEDES
                     GHVGVALSKDLMSVAAEALKTNITTLGPRVLPMSEKMKMMLVGGKAAVVQVLGVQQQG
                     QGGSRSRPYIPDLTTAFEHFCIHAGGRAVLDTVERSLDLSAWHMEPSRMTLYRWGNTS
                     SSSLWYELAYTEAKGRVRRGHRVCQIAFGSGFKCNSAVWRALRTIDPAKESAAAGNPW
                     VEEIHRFPVEVPQVERVDVAAMAS"
     repeat_region   complement(5825. .5905)
                     /rpt_family="HCSR-4"
                     /rpt_type=dispersed
     gene            complement(7346. .8203)
                     /gene="22-kDa kafirin"
                     /pseudo
     repeat_region   11416. .11598
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     mRNA            join(<11990. .12670,13053. .13158,14484. .14704,
                     14766. .>14987)
                     /product="hypothetical protein"
     CDS             join(11990. .12670,13053. .13158,14484. .14704,14766. .14987)
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD27578.1"
                     /db_xref="GI:4680215"
                     /translation="MYTPYPDSESLKTLNGQATSDHACTLNFPIQLYQIQEPESRATT
                     NVHRPLDRALLLAVTAAGVAMALHRRDVLVVHRGELRVAIPRRSDSVAPKHITEGEED
                     ESKELEVGELLGALLRGARMAVDVAVGVTRAPGGEHGTVERERRLVEPEPLQRDEDAR
                     GEVHREARRQERAQDGGGDAERHEDDEEVVVVAFETHVRRGVEVEEHQLRRERRRRYQ
                     AECAVGQREHSNAKQSQTALATVLSSSKPATHLAPLRTCKVGIKEARGGRCPAPAAAV
                     TDDRRRRRRTGRGGQTDRPGPELGGEGESAEEEEHEEHPAVLEVVEVRRVDFPAVDED
                     PESSSSRSGEEEVVVRTGFHARTRNVDGDAACRDAEHQTRHRPTSRNSVPGSTWPLND
                     PMVVRRPDGRMDRLVVD"
     repeat_region   13436. .13657
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   13994. .14148
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   complement(15740. .15838)
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   complement(15910. .16127)
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   16501. .16675
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   17050. .17249
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   complement(17650. .17893)
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   complement(19738. .19888)
                     /rpt_family="HCSR-4"
                     /rpt_type=dispersed
     repeat_region   20094. .20411
```

```
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   complement(22849. .23088)
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     repeat_region   complement(23211. .23402)
                     /rpt_family="HCSR-1"
                     /rpt_type=dispersed
     mRNA            complement(join(<26698. .27270,27361. .27761,27976. .28284,
                     29824. .31036,31131. .31286,31385. .31568,32230. .32756,
                     32998. .33453,33728. .34017,34138. .34292,34882. .>35045))
                     /product="hypothetical protein"
     CDS             complement(join(26698. .27270,27361. .27761,27976. .28284,
                     29824. .31036,31131. .31286,31385. .31568,32230. .32756,
                     32998. .33453,33728. .34017,34138. .34292,34882. .35045))
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD27577.1"
                     /db_xref="GI:4680214"
                     /translation="MDGEQAARWAAAQEGVPVGADLVAAALRQLKFLAAVDRRRWLYD
                     EGPLLHRAIRRYKACWLPLLDKHTKAAVVDGPLVVPLDCEWIWHCHRLNPVHILTMMI
                     LCNSQLSVLIIQAHIFQVQYIRDCKKVYGRILNNNNVESSTQTKSILQSEKIWKELYP
                     EEPFKLEFTKTSDVVMDVNPGVAEDITYDLVSAVKRQSSFYYQVGTPTMHDSRFLQEA
                     LARYKAFLYLIKMNQEKGLQRFRVPTYDVDLLWHTHQLHPVTYRNDMVKLLGKVLEHD
                     DTDADRSEGKKLDVGFTETTEQFENTFGVRYWKAGCMYRGNMPSPVTSTPQIFSTEVG
                     TGSDICKAQKDLNALDITAVELYLQIVDINNLPSAVRKENVYVRFTKNQPDTFISDGG
                     KLDISTVTGKNAGVCLQCEPTGELILVVMVDQVSKKPEPIGKVSFPLQDLIGPDSKLS
                     FEKWFELKAHGGHATSPPVSLRVAASATVPSSFQKVFSMVMMEPFSLKSCLLPHSIKD
                     QNMSSWTRFVYDCGTELIRLQIREQKAKNGMAAVRELVGVLKSPKKQFQLAEFKENKW
                     TLKDSMLSITHGTDGSLLDLKGDNQLIKLYQGRKLEYERKCCSAHSEDVSAVTAVKFC
                     AEHPYGKAVALLDTESQLIMVNEDWFLLPWITTSVLFMDADGKDSMKLIGGAVVQKGA
                     VFGSDTSMVLETQAVGAGSVATAPAQCGTCSTAFSGDKVVAGFKAEHASSGASRTVVV
                     SGENSHTESASGCGGGCGSSCGPMVAEDSKADNAKSGGCGSGCGGGCGGGGGCGTMLK
                     ASTMASEGQSRSGGCGSGCGGGCGGGGGCGTMLKASTMASEGQSKSGGCGGGCGSGCG
                     GGCGSGMAIEGSKTANHAKSGGCGSGCGGGCGGGGGCGTMLKASTMASVGQSKSGGCG
                     GGCGSGCGGGCGGGMAIEGSKTANHAKSGGCGSGCGSGCGGGCGTLFSSSTAADQGHS
                     RSSGGCGSGCGGGGCGSGSMVAEGSTGRHAKSGGCGSSCGGGCGGGGGCGTIFNVSTK
                     AGGEGKHSVSGGCGSGCGVLHAVLRVELSRRGISAASGEKSRADATRAARREDRDASG
                     QTRRERADGGQSGGSLAARLKIARVSILKIARVDGRRESREDRLVRPSLYLSTSPWTT
                     PSSSVGTPLFSGNVVSAQLQEDVVESQSPDVTVGGRRAKKGTSHRQKKFDIEEDKVIC
                     SAWLNVSKDPINGANQGRSTFWGRIRAFFEKHKDKKTAVCVRTESSIMHRWLTIQRDV
                     NKFCSCYEKIERRNASGATIQDMINAALETYIGADEENKPFTLMHCYLKLKEEDKWKS
                     LRIELAQKNKKQKTTKESTPSNVEASNNDEIPEVAAPDSEERKRPKGQKQVKQARRDD
                     ASLAVLEKMLEKKEARELERDKAREASLEIERASLEIEKAALEIEKKRLSNEEKKIEA
                     ELMKEEKDIMLADKTSLDEEQLQWLQIMKNKILARHTGN"
     repeat_region   36421. .36587
                     /rpt_family="HCSR-4"
                     /rpt_type=dispersed
     mRNA            complement(join(<38130. .38275,38407. .38612,38705. .38895,
                     39620. .39776,39932. .40049,40120. .40673,40764. .40864,
                     41122. .41697,41780. .42115,42344. .42727,42847. .>43083))
                     /product="hypothetical protein"
     CDS             complement(join(38130. .38275,38407. .38612,38705. .38895,
                     39620. .39776,39932. .40049,40120. .40673,40764. .40864,
                     41122. .41697,41780. .42115,42344. .42727,42847. .43083))
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD27562.1"
                     /db_xref="GI:4680199"
                     /translation="MKTSACLARVSAVEFRLPLVGLFNLLLLLQKSKANQRVQKLQQL
                     LPKSNFYCSTISKAPPPCFHRLLNDQVGENYPLPPGSSAATIGSTAAASPTATAAARR
                     RPAASQAASPVATHPTRAHGRDWLGFLRRHRCLRLDDGRLPTRGRGCSTKAGCISSRD
                     RGCISTRRLHLRLKLLRRGGRSYSDAAAASLPDDDRLPKGSRSLGRKIEIFLFGSSMG
                     DKADWSDRFLRNLFDACKEEIDAGNRPMGIFTATGWKNVVSKFGDKSGDKRTKKQLKN
                     KLDILKKEYTMFMEFKNFATGLGWDAENKTVDCPKEWWDEHLARCNNPEKGIKCSHVK
                     FRKHGPKFLEDMHVIFGKAHVDGSTATCPGDISSDDTSDEDVAEVPKPTKEKEKTVNQ
                     GKRKRKGTSTAAEEKEEKSPFFRMYKNTCLKIENAAERISTSVEVSSAPTINQIPSIA
                     EAVKMVKECGVEEGTALMYTAWSALANPEFREFFTSLETNNGRLDLIQREHEKEKNKK
                     EGSMDDHLRAVAKGQKGLLLLAELSQHAAIVGGMGNSYFYKMFRMSLEVFNELHDLLV
                     SNYGLTSTKNVSSIESLAMFLWIVGGPQSFSQAENRFTRSLWTIHTKFHEVLRCLRKL
                     AKDNITPKDPTFSTEHERVRENRFWPYFKGAIGAIDGSHVKVIVPLNETVNHTNRRGY
                     TSQNVLAICDFDMRFTFAVTGWPGSAHDSRILNHALANFPSFPVAPKGKYYLVDSGYA
                     NRNGYLAPFKGSTYHIPEFRNRSEPPQGNKMRDKEFDRCDADEDYMPRCARRRTQTHE
                     DGTDEGESDESMNIIRDRIVDALTKEKTRARRRAEDSLNTRGGGGRSVPQAVRSHGAL
                     RRPPLVVSGGGRRRRRDGAVVQSSHVQPRICTAHHLQQEPRLVQGQGIQQVAVHQVRA
                     AEREAGRRRRTRRVAGPRGHGRLRLRDRRRGRHRQRLPPELRRGYSGADAGARRVGAR
                     RRPRRLLPAPVRRRLARARLMNTHRHIAM"
     mRNA            join(<43647. .43908,43994. .44222,44944. .>45484)
                     /product="hypothetical protein"
     CDS             join(43647. .43908,43994. .44222,44944. .45484)
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD27563.1"
                     /db_xref="GI:4680200"
                     /translation="MSGGIQDDSDLIWLSVRMIPRYKQVLSFLFGWSHGIRCQHPSID
                     LVRLPQFSREQTREQGYRTNDLFVEEQHCRRIASTSDVCTAQEAAQRSADAVGLHPPR
                     LSLDLLNISETFAFVASLLHPESRAAAALATARSSCSLAAASGSTAQSASSTLKRWPSW
                     IDRATSFTTTFTVYIERLTGMEIHCKDVQDQEIRSLLRPQPDARVAVNFKIDEVYPSR
                     DNAVRLSDPNIQTNIESPSPVFLVQIAPRPPPPLSPVCSRNLCRPGAHVACALCVWRG
                     RCRCRRRWDNTTGRQMLIFMPADYYTHVTTVIIYDDDDSIIHALLLVLLACGQLLALC
                     TLVYHLRCT"
     repeat_region   complement(45613. .45755)
                     /rpt_family="HCSR-4"
                     /rpt_type=dispersed
     mRNA            complement(join(<46323. .46753,47300. .47458,48167. .>48248))
                     /product="hypothetical protein"
     CDS             complement(join(46323. .46753,47300. .47458,48167. .48248))
                     /note="similar to Oryza sativa EST clone C3025_1A"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD27561.1"
                     /db_xref="GI:4680198"
```

```
  Query Match            92.8%;  Score 389.8;  DB 8;  Length 183990;
  Best Local Similarity  95.9%;  Pred. No. 2.9e-99;
  Matches  400;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101567 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 101508

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||| |||||||||||||||| ||||||||||||
Db  101507 ACTAAAGCAAGTGTTGGATTTAAAGCTGATGTTAAGGATTATAAATTAACTTACTACACC 101448

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           |||||||| |||| ||||||||| ||||||||||||||||||||||||||||||||||
Db  101447 CCGGAGTACAAAACGAAGGATACTAATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 101388

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||
Db  101387 GGGGTTCCGCCTGAAGAAGTAGGAGCTGCAGTAGCTGCGGAATCTTCTACTCGTACATGG 101328

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           |||||| |||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db  101327 ACAACTATTTGGACTGATGGACTTACCAGTCTTGATTGTTACAAAGGATGATGCTATCAC 101268

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  101267 ATCGAGCCCATTCCTAGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 101208

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||| | ||||||||||||||||| |||||||||||||||| ||
Db  101207 CTATTTGAAGAGGGTTTTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTACTT 101151
```

```
RESULT 6
CHZM02
LOCUS       CHZM02                  1803 bp    DNA     linear   PLN 29-MAR-2001
DEFINITION  Zea mays chloroplast gene for the large subunit of RUBP (ribulose
            bisphosphate carboxylase).
ACCESSION   V00171 J01423
VERSION     V00171.1  GI:12394
KEYWORDS    ribulose bisphosphate carboxylase.
SOURCE      Zea mays.
  ORGANISM  Chloroplast Zea mays
```

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1803)
  AUTHORS   McIntosh,L., Poulsen,C. and Bogorad,L.
  TITLE     Chloroplast gene sequence for the large subunit of ribulose
            bisphosphatecarboxylase of maize
  JOURNAL   Nature 288, 556-560 (1980)
COMMENT     On Jul 24, 2002 this sequence version replaced gi:342623.
FEATURES             Location/Qualifiers
     source          1. .1803
                     /organism="Zea mays"
                     /organelle="plastid:chloroplast"
                     /db_xref="taxon:4577"
     CDS             228. .1655
                     /codon_start=1
                     /transl_table=11
                     /product="ribulose bisphosphate carboxylase"
                     /protein_id="CAA23474.1"
                     /db_xref="GI:12395"
                     /db_xref="SWISS-PROT:P00874"
                     /translation="MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVTP
                     QLGVPPEEAGAAVAAESSAAGTWTTVWTDGLTSLDRYKGRCYHIEPVPGDPDQYICYV
                     AYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYSKTFQGPPRGMQVER
                     DKLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDENVNSQPFMRWRDR
                     FVFCAEAIYKSQAETGEIKGHYLNATAGTCDEMIKGAVFARQLGVPIVMHDYLTGGFT
                     ANTTLSHYCRDNGLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHSGTVVGK
                     LEGEREITLGFVDLLRDDFIEKDRSRGIFFTQDWVSMPGVIPVASGGIHVWHMPALTE
                     ILGDDSVLQFGGGTLGHPWGNAHGAAANRVALEACVQARNEGRDLAREVQIIKAACKW
                     SAELAAACEIWKEIKFDGFKAMDTI"
BASE COUNT      539 a    327 c    410 g    527 t
ORIGIN

  Query Match             91.9%;  Score 386;  DB 8;  Length 1803;
  Best Local Similarity   96.9%;  Pred. No. 2.8e-98;
  Matches  405;  Conservative    0;  Mismatches   10;  Indels    3;  Gaps    1;

Qy      6 TTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAAC 65
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188 TTGTGTGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAAC 247

Qy     66 TAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCC 125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248 TAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCC 307

Qy    126 GGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGG 185
          |||||||||||||||||||||||||||||| |||||||||||||||| ||||||||
Db    308 GGAGTACGAAACCAAGGATACTGATATCTTGGCGGCATTCCGAGTAACTCCACAGCTCGG 367

Qy    186 GGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTT---CTACTGGTACATG 242
          |||||||||||| |||||||| |||||||| ||||||||||||   || ||||||||||
Db    368 GGTTCCGCCTGAGGAAGCAGGGGCTGCAGTGGCTGCGGAATCTTCTGCTGCTGGTACATG 427

Qy    243 GACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCA 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    428 GACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCA 487

Qy    303 CATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGA 362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    488 CATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGA 547

Qy    363 CCTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db    548 CCTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 605


RESULT 7
CHOSRBCL
LOCUS       CHOSRBCL                1989 bp    DNA     linear   PLN 29-MAR-2001
DEFINITION  Rice chloroplast DNA for ribulose-1,5-biphosphate carboxylase
            (clone pCT-1).
ACCESSION   X04789
VERSION     X04789.1  GI:11954
KEYWORDS    ribulose bisphosphate carboxylase.
SOURCE      Oryza sativa.
  ORGANISM  Plastid Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1989)
  AUTHORS   Moon,E., Kao,T.H. and Wu,R.
  TITLE     Rice chloroplast DNA molecules are heterogeneous as revealed by DNA
            sequences of a cluster of genes
  JOURNAL   Nucleic Acids Res. 15 (2), 611-630 (1987)
  MEDLINE   87146381
   PUBMED   3029686
FEATURES             Location/Qualifiers
     source          1. .1989
                     /organism="Oryza sativa"
                     /organelle="plastid"
                     /strain="Labelle"
                     /db_xref="taxon:4530"
                     /clone="pCt-1"
     promoter        41. .46
                     /note="pot. -35 region"
     promoter        65. .70
                     /note="pot. -10 region"
     RBS             377. .383
                     /note="pot. ribosome binding site"
     CDS             388. .1821
                     /note="ribulose-1,5-bisphosphate carboxylase (AA 1 - 477)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA28475.1"
                     /db_xref="GI:11955"
                     /db_xref="SPTREMBL:Q32753"
                     /translation="MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVTP
                     QPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVVGEDNQYIAYVA
                     YPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPTYSKTFQGPPHGMEVERD
                     KLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDENVNSQPFMRWRDRF
                     VFCAEAIYKSQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTA
                     NTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGAVVGK
                     LEGEREMTLGFVDLLRDDFIEKDRARGIFFTQDWVSMPGVIPVASGGIHVWHMPALTE
                     IFGDDSVLQFGGGTLGHPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIRSACK
                     WSPELAAACEIWKAIKFEFEPVDKLDS"
BASE COUNT      600 a    348 c    452 g    589 t
ORIGIN

  Query Match             91.7%;  Score 385;  DB 8;  Length 1989;
  Best Local Similarity   95.2%;  Pred. No. 5.4e-98;
  Matches  397;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    346 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 405

Qy     64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 465

Qy    124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    466 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 525

Qy    184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db    526 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 585

Qy    244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    586 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 645

Qy    304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||  ||||||     |||||||||||  ||||||||||||||||||||||
Db    646 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 705
```

```
Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           ||||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     706 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 762


RESULT 8
RICCPRBCL
LOCUS       RICCPRBCL               2062 bp    DNA     linear   PLN 28-APR-2001
DEFINITION  Oryza sativa chloroplast gene for ribulose 1,5-bisphosphate
            carboxylase large subunit, complete cds.
ACCESSION   D00207
VERSION     D00207.1  GI:344016
KEYWORDS    Rubisco LS; chloroplast; large subunit; rbcL; ribulose 1,
            5-bisphosphate carboxylase.
SOURCE      Oryza sativa chloroplast DNA.
  ORGANISM  Chloroplast Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2062)
  AUTHORS   Nishizawa,Y. and Hirai,A.
  TITLE     Nucleotide sequence and expression of the gene for the large
            subunit of rice ribulose 1, 5-bisphosphate carboxylase
  JOURNAL   Jpn. J. Genet. 62, 389-395 (1987)
COMMENT     The initiation and termination sites of the Rubisco LS mRNA were
            also determined by S1 mapping method.
FEATURES             Location/Qualifiers
     source          1. .2062
                     /organism="Oryza sativa"
                     /organelle="plastid:chloroplast"
                     /db_xref="taxon:4530"
                     /note="588 bp upstream of PstI site"
     -35_signal      74. .79
     -10_signal      98. .103
     mRNA            108. .>1854
     RBS             411. .415
     gene            421. .1854
                     /gene="rbcL"
     CDS             421. .1854
                     /gene="rbcL"
                     /note="Rubisco LS"
                     /codon_start=1
                     /transl_table=11
                     /product="ribulose 1,5-bisphosphate carboxylase large
                     subunit"
                     /protein_id="BAA00147.1"
                     /db_xref="GI:344017"
                     /translation="MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVTP
                     QPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVVGEDNQYIAYVA
                     YPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPTYSKTFQGPPHGIQVERD
                     KLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDENVNSQPFMRWRDRF
                     VFCAEAIYKSQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTA
                     NTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGK
                     LEGEREMTLGFVDLLRDDFIEKDRARGIFFTQDWVSMPGVIPVASGGIHVWHMPALTE
                     IFGDDSVLQFGGGTLGHPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIRSACK
                     WSPELAAACEIWKAIKFEFEPVDKLDS"
BASE COUNT      626 a    358 c    468 g    610 t
ORIGIN

  Query Match           91.7%;  Score 385;  DB 8;  Length 2062;
  Best Local Similarity   95.2%;  Pred. No. 5.4e-98;
  Matches  397;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     379 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 438

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 498

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||

Db     499 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 558

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db     559 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 618

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     619 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 678

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||| ||||||  ||||||||||| ||||||||||||||||||||||||||
Db     679 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 738

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           ||||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     739 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 795


RESULT 9
AC122148
LOCUS       AC122148               125936 bp    DNA     linear   PLN 06-JUN-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 10 clone
            OSJNAb0075K12, complete sequence.
ACCESSION   AC122148
VERSION     AC122148.1  GI:21104881
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 125936)
  AUTHORS   Wing,R.A., Yu,Y., Yang,T.J., Nah,G., Soderlund,C., Chen,M.,
            Kim,H.-R., Rambo,T., Saski,C., Henry,D., Oates,R. and Simmons,J.
  TITLE     Rice Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 125936)
  AUTHORS   Wing,R.A., Yu,Y., Yang,T.J., Nah,G., Soderlund,C., Chen,M.,
            Kim,H.-R., Rambo,T., Saski,C., Henry,D., Oates,R. and Simmons,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
REFERENCE   3  (bases 1 to 125936)
  AUTHORS   Wing,R.A., Yu,Y., Yang,T.J., Nah,G., Soderlund,C., Chen,M., Kim,H.,
            Rambo,T., Saski,C., Henry,D., Oates,R., Simmons,J. and
            Thurmond,S.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
COMMENT     The following sequence is an artificial clone that will be used for
            gap filling in the rice genome.  There is 60 kb of sequence from
            the overlapping north clone (OSJNBb0075K12) and 60 kb of sequence
            from the overlapping south clone (OSJNBa0022D10) and the gap
            filling sequence begins at 60,001 and ends at 65,936 for a total
            size of 125,936 bases.  The gap filling sequence is 5936 bases. The
            name (OSJNAb0075K12) was derived from the location of the north
            clone and the 'A' stands for Artificial.
FEATURES             Location/Qualifiers
     source          1. .125936
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /db_xref="taxon:39947"
                     /chromosome="10"
                     /clone="OSJNAb0075K12"
     gene            complement(<1. .1590)
                     /gene="OSJNAb0075K12.1"
                     /note="3' Partial - Hypothetical protein ORF542 from
                     chromosome 10 chloroplast insertion"
     misc_feature    complement(<1. .1590)
                     /gene="OSJNAb0075K12.1"
                     /note="3' Partial - Hypothetical protein ORF542 from
                     chromosome 10 chloroplast insertion"
     repeat_region   complement(2437. .2462)
```

```
                /note="Low_complexity AT_rich"
gene            complement(2781. .3848)
                /gene="OSJNAb0075K12.2"
                /note="Putative ribosomal protein S16 from chromosome 10
                chloroplast insertion"
CDS             complement(join(2781. .2929,3809. .3848))
                /gene="OSJNAb0075K12.2"
                /codon_start=1
                /product="Putative ribosomal protein S16 from chromosome
                10 chloroplast insertion"
                /protein_id="AAM48248.1"
                /db_xref="GI:21327343"
                /translation="MLKLRLKRCGRKQRFYDPIKNQTCLNVPAILYFLEKGAQPTRTV
                SDILRKAEFFKEKERTLS"
repeat_region   complement(4214. .4238)
                /note="Low_complexity AT_rich"
repeat_region   complement(4610. .4644)
                /note="Low_complexity AT_rich"
gene            5327. .5512
                /gene="OSJNAb0075K12.3"
                /note="Putative PSII K protein from chromosome 10
                chloroplast insertion"
CDS             5327. .5512
                /gene="OSJNAb0075K12.3"
                /codon_start=1
                /product="Putative PSII K protein from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48249.1"
                /db_xref="GI:21327344"
                /translation="MPNILSLTCICFNSVIYPTSFFFAKLPEAYAVFNPVVDFMPVIP
                VLFFLLAFIWQAAVSFR"
gene            5901. .6011
                /gene="OSJNAb0075K12.4"
                /note="Putative PSII I protein from chromosome 10
                chloroplast insertion"
CDS             5901. .6011
                /gene="OSJNAb0075K12.4"
                /codon_start=1
                /product="Putative PSII I protein from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48250.1"
                /db_xref="GI:21327345"
                /translation="MLTLKLFVYTVVILFVSLFIFGFLSNDPGRNPGRDE"
gene            6642. .6875
                /gene="OSJNAb0075K12.5"
                /note="Hypothetical protein ORF100 from chromosome 10
                chloroplast insertion"
CDS             6642. .6875
                /gene="OSJNAb0075K12.5"
                /codon_start=1
                /product="Hypothetical protein ORF100 from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48251.1"
                /db_xref="GI:21327346"
                /translation="MKLYKWIFLYFLRLRPRKARISRYLLFHILVTRCSIGIRRNRNP
                FPLHIKMDFESQFFFSISLLFFIFPLKNRLSWK"
gene            8133. .9554
                /gene="OSJNAb0075K12.6"
                /note="Putative PSII 43kDa protein from chromosome 10
                chloroplast insertion"
CDS             8133. .9554
                /gene="OSJNAb0075K12.6"
                /codon_start=1
                /product="Putative PSII 43kDa protein from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48252.1"
                /db_xref="GI:21327347"
                /translation="MKILYSLRRFYHVEMLFNGTFVLAGRDQETTGFAWWAGNARLIN
                LSGKLLGAHVAHAGLIVFWAGAMNLFEVAHFVPEKPMYEQGLILLPHLATLGWGVGPG
                GEVLDTFPYFVSGVLHLISSAVLGFGGIYHALLGPETLEESFPFFGYVWKDRNKMTTI
                LGIHLILLGIGAFLLVLKALYFGGIYDTWAPGGGDVRKITNLTLSPGVIFGYLLKSPF
                GGEGWIVSVDDLEDIIGGHVWLGFICVFGGIWHILTKPFAWAHRAFVWSGEAYLSYSL
                GALSVFGFIACCFVWFNNTAYSSEFYGPTGPEASQAQAFTFLVRDQRLGANVGSAQGP
                TGLGKYLMRSPTGEVIFGGETMRFWDLRAPWLEPLRGPNGLDLSRLKKDIQPWQERHS
                AEYMTHAPLGSLNSVGGVATEINAVNYVSPRSWLATSHFVLGFFFFVGHLWHAGRARA
                AAAGFEKGIDRDLEPVLYMTPLN"
gene            10161. .10349
                /gene="OSJNAb0075K12.7"
                /note="Hypothetical protein ORF62 from chromosome 10
                chloroplast insertion"
CDS             10161. .10349
                /gene="OSJNAb0075K12.7"
                /codon_start=1
                /product="Hypothetical protein ORF62 from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48253.1"
                /db_xref="GI:21327348"
                /translation="MTIAFQLAVFALIVTSSVLVISVPFVFASPDGWSNNKNVVFSGT
                SLWIGLVFLVAILNSLIS"
gene            complement(12297. .12566)
                /gene="OSJNAb0075K12.8"
                /note="Hypothetical protein ORF91 from chromosome 10
                chloroplast insertion"
CDS             complement(12297. .12566)
                /gene="OSJNAb0075K12.8"
                /codon_start=1
                /product="Hypothetical protein ORF91 from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48254.1"
                /db_xref="GI:21327349"
                /translation="MVYGYGKSNMPHPNRKRKGTDTQYDYWEELLVMVSGLYALFCVF
                LVLFIFFDSFKQESNKLELSGKEEKKKLNGENRLSRDIQNLLYIK"
gene            complement(12791. .13003)
                /gene="OSJNAb0075K12.9"
                /note="Hypothetical protein ORF70 from chromosome 10
                chloroplast insertion"
CDS             complement(12791. .13003)
                /gene="OSJNAb0075K12.9"
                /codon_start=1
                /product="Hypothetical protein ORF70 from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48255.1"
                /db_xref="GI:21327350"
                /translation="MKTFSIYVLFPLIGWGINEFAPFIYPYGWVLIFSDILLVCFGER
                LKQFLTGIPWKNLRVQNFAIFLKRIC"
gene            14907. .15011
                /gene="OSJNAb0075K12.10"
                /note="Putative PSII low MW protein from chromosome 10
                chloroplast insertion"
CDS             14907. .15011
                /gene="OSJNAb0075K12.10"
                /codon_start=1
                /product="Putative PSII low MW protein from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48256.1"
                /db_xref="GI:21327351"
                /translation="MEVNILAFIATALFILVPTAFLLIIYVKTVSQND"
gene            complement(15774. .15863)
                /gene="OSJNAb0075K12.11"
                /note="Hypothetical protein ORF29 from chromosome 10
                chloroplast insertion"
CDS             complement(15774. .15863)
                /gene="OSJNAb0075K12.11"
                /codon_start=1
                /product="Hypothetical protein ORF29 from chromosome 10
                chloroplast insertion"
                /protein_id="AAM48257.1"
                /db_xref="GI:21327352"
                /translation="MDIVSLAWAALMVVFTFSLSLVVWGRSGL"
repeat_region   17155. .17317
                /note="Similar to Oryza sativa telomere-associated
                sequence"
repeat_region   complement(20592. .20717)
                /note="Similar to Oryza sativa telomere-associated
                sequence"
repeat_region   complement(27558. .27579)
```

```
                     /note="Low_complexity AT_rich"
     gene            27754. .28464

  Query Match             91.7%;  Score 385;  DB 8;  Length 125936;
  Best Local Similarity   95.2%;  Pred. No. 6.4e-98;
  Matches  397;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   52263 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 52322

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52323 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 52382

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   52383 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 52442

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db   52443 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 52502

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   52503 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 52562

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          |||||||||||  ||||||     |||||||||||  ||||||||||||||||||||||
Db   52563 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 52622

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db   52623 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 52679
```

RESULT 10
CHOSXX
LOCUS CHOSXX 134525 bp DNA circular PLN 29-MAR-2001
DEFINITION Oryza sativa complete chloroplast genome.
ACCESSION X15901
VERSION X15901.1 GI:11957
KEYWORDS 16S ribosomal RNA; 23S ribosomal RNA; 4.5S ribosomal RNA; atpA gene; ATPase; ATPase alpha; ATPase beta; ATPase epsilon; atpB gene; atpE gene; atpF gene; atpH gene; atpI gene; circular; circular genome; complete genome; cytochrome; cytochrome b/f complex; cytochrome b559; cytochrome B6; cytochrome f; infA gene; initiation factor; NADH dehydrogenase; NADH dehydrogenase subunit 1; NADH dehydrogenase subunit 2; NADH dehydrogenase subunit 3; NADH dehydrogenase subunit 4; NADH dehydrogenase subunit 4L; NADH dehydrogenase subunit 5; NADH dehydrogenase subunit 6; ndhA gene; ndhB gene; ndhC gene; ndhD gene; ndhE gene; ndhF gene; ndhG gene; ndhH gene; petA gene; petB gene; petD gene; petE gene; photosynthesis; photosystem I; photosystem II; psaA gene; psaB gene; psaC gene; psbA gene; psbB gene; psbC gene; psbD gene; psbE gene; psbF gene; psbG gene; psbH gene; psbI gene; psbK gene; psbL gene; psbM gene; psbN gene; pseudogene; rbcL gene; ribosomal protein; ribosomal protein L14; ribosomal protein L16; ribosomal protein L2; ribosomal protein L20; ribosomal protein L22; ribosomal protein L23; ribosomal protein L33; ribosomal protein L36; ribosomal protein S11; ribosomal protein S12; ribosomal protein S14; ribosomal protein S15; ribosomal protein S16; ribosomal protein S18; ribosomal protein S19; ribosomal protein S3; ribosomal protein S4; ribosomal protein S7; ribosomal protein S8; ribulose bisphosphate carboxylase; RNA polymerase; RNA polymerase alpha-subunit; RNA polymerase beta subunit; rpoB gene; rpoC1 gene; rpoC2 gene; transfer RNA-Ala; transfer RNA-Arg; transfer RNA-Asp; transfer RNA-Cys; transfer RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His; transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer RNA-Met; transfer RNA-Phe; transfer RNA-Ser; transfer RNA-Thr; transfer RNA-Tyr; transfer RNA-Val.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Plastid Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 134525)
AUTHORS Hiratsuka,J., Shimada,H., Whittier,R., Ishibashi,T., Sakamoto,M., Mori,M., Kondo,C., Honji,Y., Sun,C.R., Meng,B.Y., Li,Y.Q., Kanno,A., Nishizawa,Y., Hirai,A., Shinozaki,K. and Sugiura,M.
TITLE The complete sequence of the rice (Oryza sativa) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals
JOURNAL Mol. Gen. Genet. 217 (2-3), 185-194 (1989)
MEDLINE 89364698
PUBMED 2770692
REFERENCE 2 (bases 1 to 134525)
AUTHORS Shimada,H. and Sugiura,M.
TITLE Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes
JOURNAL Nucleic Acids Res. 19 (5), 983-995 (1991)
MEDLINE 91212240
PUBMED 1708498
REFERENCE 3 (bases 1 to 134525)
AUTHORS Morton,B.R. and Clegg,M.T.
TITLE A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near rbcL in the grass family (Poaceae)
JOURNAL Curr. Genet. 24 (4), 357-365 (1993)
MEDLINE 94073993
PUBMED 8252646
REFERENCE 4 (bases 1 to 134525)
AUTHORS Sugiura,M.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-1989) Sugiura M., Centre for Gene Research, Nagoya University, Chikusa Nagoya 464-01, Japan

```
FEATURES             Location/Qualifiers
     source          1. .134525
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /organelle="plastid"
                     /cultivar="Nipponbare."
                     /db_xref="taxon:39947"
     CDS             complement(join(88473. .88501,89042. .89273,67390. .67503))
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein S12"
                     /protein_id="CAA33929.1"
                     /db_xref="GI:1213603"
                     /db_xref="SWISS-PROT:P12149"
                     /translation="MPTVKQLIRNARQPIRNARKSAALKGCPQRRGTCARVYTINPKK
                     PNSALRKVARVRLTSGFEITAYIPGIGHNLQEHSVVLVRGGRVKDLPGVRYRIIRGAL
                     DAVAVKNRQQGRSKYGVKKPKK"
     gene            complement(82. .1143)
                     /gene="psbA"
     CDS             complement(82. .1143)
                     /gene="psbA"
                     /codon_start=1
                     /transl_table=11
                     /product="PSII 32kDa protein"
                     /protein_id="CAA34007.1"
                     /db_xref="GI:11958"
                     /db_xref="SWISS-PROT:P12094"
                     /translation="MTAILERRESTSLWGRFCNWITSTENRLYIGWFGVLMIPTLLTA
                     TSVFIIAFIAAPPVDIDGIREPVSGSLLYGNNIISGAIIPTSAAIGLHFYPIWEAASV
                     DEWLYNGGPYELIVLHFLLGVACYMGREWELSFRLGMRPWIAVAYSAPVAAATAVFLI
                     YPIGQGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHMLGVAGVFGGSLFSAMHGSL
                     VTSSLIRETTENESANEGYRFGQEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLA
                     AWPVVGIWFTALGISTMAFNLNGFNFNQSVVDSQGRVINTWADIINRANLGMEVMHER
                     NAHNFPLDLAALEVPSLNG"
     tRNA            complement(1363. .1397)
                     /product="tRNA-Lys"
                     /note="trnK ;tRNA-Lys(UUU) 3'exon"
     CDS             complement(1668. .3296)
                     /note="ORF542; ORF within trnK intron"
                     /codon_start=1
                     /transl_table=11
```

```
                /protein_id="CAA34008.1"
                /db_xref="GI:11959"
                /db_xref="SWISS-PROT:P12175"
                /translation="MQHPVLTILYYVSPFDKPRNISLSLIQVEIQMEKFEGYSEKLKF
                PRQYFVYPLLFQEYIYVFAHDYGLNGSELVEIIGSNNKKFSSLLVKRLMIRMYQQNFW
                INLVNHPNQDRLLDYNNFFYSEFYSQILSEGFAIVVEIPFSLREQSCPEEKEIPKFQN
                LRSIHSIFPFLEDKFLHLHYLAHIEIPYPIHLDILLQLLQYRIQDVPSLHLLRFFLNY
                YSNWNSFITSMKSIFILKKENKRLFRFLYNSYVSEYEFFLLFLRKQSSCLRLTSSGTF
                LERIIFSRKMEHFGLMYPAFFRKTIWFVMDPLMHYVRYQGKAILASKGTLLLKKKWKC
                YLVRLWQYSFSFWTQPQRIHLNQLENSCFDFLGYFSSVPINSLLVRNQMLENSFLIDT
                QMKKFDTKVPVTPLIGSLAKAQFCTGSGHPISKPIWTDLSDWDILDRFGRICRNLFHY
                HSGSSKKKTLYRLKYILRLSCARTLARKHKSTVRAFMQWLGSVFLEEFFTEEEQVFSL
                MFAKTTYFSFRGSHSERIWYLDILRINDLVNPLN"
tRNA            complement(3902. .3938)
                /product="tRNA-Lys"
                /note="trnK ;tRNA-Lys(UUU) 5'exon"
gene            complement(4487. .5553)
                /gene="rps16"
CDS             complement(join(4487. .4635,5514. .5553))
                /gene="rps16"
                /codon_start=1
                /transl_table=11
                /product="ribosomal protein S16"
                /protein_id="CAA34009.1"
                /db_xref="GI:669079"
                /db_xref="SWISS-PROT:P12151"
                /translation="MLKLRLKRCGRKQRFYDPIKNQTCLNVPAILYFLEKGAQPTRTV
                SDILRKAEFFKEKERTLS"
exon            complement(4487. .4635)
                /gene="rps16"
                /number=2
exon            complement(5514. .5553)
                /gene="rps16"
                /number=1
tRNA            complement(6615. .6687)
                /product="tRNA-Gln"
                /note="trnQ; tRNA-Gln(UUG)"
gene            7033. .7218
                /gene="psbK"
CDS             7033. .7218
                /gene="psbK"
                /codon_start=1
                /transl_table=11
                /product="PSII K protein"
                /protein_id="CAA34010.1"
                /db_xref="GI:11961"
                /db_xref="SWISS-PROT:P12162"
                /translation="MPNILSLTCICFNSVIYPTSFFFAKLPEAYAIFNPIVDFMPVIP
                VLFFLLAFVWQAAVSFR"
gene            7608. .7718
                /gene="psbI"
CDS             7608. .7718
                /gene="psbI"
                /codon_start=1
                /transl_table=11
                /product="PSII I protein"
                /protein_id="CAA34011.1"
                /db_xref="GI:11962"
                /db_xref="SWISS-PROT:P12161"
                /translation="MLTLKLFVYTVVIFFVSLFIFGFLSNDPGRNPGRDE"
tRNA            complement(7829. .7916)
                /product="tRNA-Ser"
                /note="trnS; tRNA-Ser(GCU)"
CDS             8349. .8651
                /note="ORF100"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA34012.1"
                /db_xref="GI:11963"
                /db_xref="SPTREMBL:Q32758"
                /translation="MKLYKWIFLYFLRLRPRKARISRYLLFHILVTRCSIGIRRNRNP
                FPLHIKMDFESQFFFSISLLFFRILFLFLYPCNRERVGKGRLLFFIFSLKNRLSWK"
gene            8900. .9961
                /gene="psbD"
CDS             8900. .9961
                /gene="psbD"
                /codon_start=1
                /transl_table=11
                /product="PSII D2 protein"
                /protein_id="CAA34013.1"
                /db_xref="GI:11964"
                /db_xref="SWISS-PROT:P12095"
                /translation="MTIALGRVTKEENDLFDIMDDWLRRDRFVFVGWSGLLLFPCAYF
                ALGGWFTGTTFVTSWYTHGLASSYLEGCNFLTAAVSTPANSLAHSLLLLWGPEAQGDF
                TRWCQLGGLWTFVALHGAFALIGFMLRQFELARSVQLRPYNAISFSGPIAVFVSVFLI
                YPLGQSGWFFAPSFGVAAIFRFILFFQGFHNWTLNPFHMMGVAGVLGAALLCAIHGAT
                VENTLFEDGDGANTFRAFNPTQAEETYSMVTANRFWSQIFGVAFSNKRWLHFFMLFVP
                VTGLWMSAIGVVGLALNLRAYDFVSQEIRAAEDPEFETFYTKNILLNEGIRAWMAAQD
                QPHENLIFPEEVLPRGNAL"
gene            9909. .11330
                /gene="psbC"
CDS             9909. .11330
```

Query Match 91.7%; Score 385; DB 8; Length 134525;
Best Local Similarity 95.2%; Pred. No. 6.4e-98;
Matches 397; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   54053 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 54112

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54113 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 54172

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   54173 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 54232

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db   54233 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 54292

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   54293 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 54352

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          |||||||||||| |||||| |||||||||||| |||||||||||||||||||||||||||
Db   54353 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 54412

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| |||||||||||||||||||||
Db   54413 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 54469
```

```
RESULT 11
AC092750
LOCUS       AC092750              134933 bp    DNA     linear   PLN 04-APR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) cultivar Nipponbare from
            chromosome 10, complete sequence.
ACCESSION   AC092750
VERSION     AC092750.2  GI:16356860
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 134933)
  AUTHORS   McCombie,W.R., de la Bastide,M., Spiegel,L., Nascimento,L.,
            Zutavern,T., Balija,V., Bell,M., Preston,R., Kirchoff,K., Kuit,K.,
            Baker,J., Santos,L., Miller,B., Cunnius,D.M., Katzenberger,F.,
            Muller,S., Shah,R., King,L., Yang,C., Dike,S., O'Shaughnessy,A.,
            Palmer,L. and Dedhia,N.
  TITLE     Genomic sequence for Oryza sativa, Nipponbare strain, clone
```

OSJNBb0075K12, from chromosome 10, complete sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 134933)
AUTHORS McCombie,W.R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY 11724, USA
REFERENCE 3 (bases 1 to 134933)
AUTHORS McCombie,W.R.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY 11724, USA
REFERENCE 4 (bases 1 to 134933)
AUTHORS Palmer,L.E., Yu,M., de la Bastide,M., Spiegel,L., Nascimento,L., Zutavern,T., Balija,V., Bell,M., Preston,R., Kirchoff,K., Kuit,K., Baker,J., Santos,L., Miller,B., Cunnius,D.M., Katzenberger,F., Muller,S., Shah,R., King,L., Yang,C., Dike,S., O'Shaughnessy,A., Dedhia,N. and McCombie,W.R.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
REMARK Genomic sequence for Oryza sativa (japonica cultivar-group) cultivar Nipponbare clone OSJNBb0075K12, from chromosome 10
COMMENT On Oct 24, 2001 this sequence version replaced gi:15011683.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES Location/Qualifiers

source 1. .134933
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="10"
/clone="OSJNBb0075K12"
/clone_lib="EcoRI"
/note="japonica cultivar-group"

repeat_region complement(20. .226)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

gene <124. .2469
/gene="OSJNBb0075K12.1"
/note="Putative retroelement"

CDS join(<124. .958,1238. .1497,1930. .2469)
/gene="OSJNBb0075K12.1"
/note="Similar to RETROVIRUS-RELATED POL POLYPROTEIN FROM TRANSPOSON 297 [CONTAINS"
/codon_start=1
/product="Putative retroelement"
/protein_id="AAM08567.1"
/db_xref="GI:19920135"
/translation="QHKLFAKPSKCSFALQQLEYLGHIISKTRVATDPAKTLVMQNWP
VPSLVADLRGFLGLTGYYRKFVKYYGLLAKPLIELLKKNLPFQWTSVTQQAFETLKQA
MITTPVLSLPDFSKPFVIETDTCDTGIGDVLSQEGHPIAYYSKALGTANQKLSIYEKE
FMAIMMAVDRWRPYLLRGPFVIKTDHKFLCHLDTHVLDSEMQKKAMTKLIGLQYHFQY
KKGVDNKSADALSRVGQFFACSSVSLVQPPWIQEIVNSYAVDMQAQKLLQELAVISPN
DQGVWQDISLDFIDGLPLSDSKNVILVVVDRFTKFAHFIPLKHPYTAQHVASVFLDRV
ASIYELPKTMVSDRVAIFTSQFWKVLFAKMKKSVASRPYPKLAFKYFGPFEILSKVGS
VAYRLKLPVDSQVHPVFHVSQLKHFVPSHVPVFSELPTTVQLDISALEPTEILDRRLV
KKGNAALTQVLVRLGVLPPECATWEDYSVLRAHFPLATVWGHPVSHGGGSVTPVGVAP
EDAGITQQRYREDGLRKRRLPVKKAYLCNKCVCARA"

repeat_region 406. .690
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region complement(1303. .1562)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region complement(4245. .4320)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region complement(5779. .5957)
/note="Simple_repeat (CCGGG)n"

repeat_region 6170. .6328
/note="Simple_repeat (CCGGG)n"

repeat_region complement(6495. .6547)
/note="Low_complexity AT_rich"

repeat_region 7667. .9036
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region complement(9725. .9807)
/note="Similar to unknown repetitive sequence"

repeat_region complement(9908. .10030)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region complement(10048. .10105)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region complement(10336. .10423)
/note="Similar to unknown repetitive sequence"

repeat_region 10468. .10552
/note="Similar to unknown repetitive sequence"

repeat_region complement(10568. .10643)
/note="Similar to unknown repetitive sequence"

gene complement(10698. .12424)
/gene="OSJNBb0075K12.2"
/note="Hypothetical protein"

CDS complement(join(10698. .11211,11988. .12424))
/gene="OSJNBb0075K12.2"
/note="Similar to (AP002525) hypothetical protein similar to Oryza sativa chromosome 6, P0680A03.17 E-value=e-105 Identities=210/732 (of subject)"
/codon_start=1
/product="Hypothetical protein"
/protein_id="AAM08568.1"
/db_xref="GI:19920136"
/translation="MDSRRLSFPVDQPRVHSAGLSSPPPLLIATTTSPPACLRRRRLV
SPPTRLACYAAVVGRLRLHRRLASVAADWCHRLHGWPIMPLLLGVYNFIAGLPPSPPT
SVLAYMVGWSHHRLWASASSPTDWPSPPPTGVSAYMASLLCRRWIQILIGTKGIHGFK
LISYALGMDQSEHHHRLAVVAADWCFRLHGWPVMPPLLGVYVFTDRLALVRRRLVFPP
ARLAYYAAVVGRLRLHRPAASSAADWCFCLHGWPIMPPLLGVYDFTDRPPRPPPTGVS
ACTTGLLCRLCWASTSSPTGRLVRRQLVFPPARLAYYDAN"

repeat_region 10704. .10778
/note="Similar to unknown repetitive sequence"

repeat_region 10779. .10867
/note="Similar to unknown repetitive sequence"

repeat_region 10868. .10956
/note="Similar to unknown repetitive sequence"

repeat_region 10957. .11047
/note="Similar to unknown repetitive sequence"

repeat_region complement(11048. .11133)
/note="Similar to unknown repetitive sequence"

repeat_region complement(11932. .11976)
/note="Similar to unknown repetitive sequence"

repeat_region 11978. .12066
/note="Similar to unknown repetitive sequence"

repeat_region 12069. .12156
/note="Similar to unknown repetitive sequence"

repeat_region 12157. .12245
/note="Similar to unknown repetitive sequence"

repeat_region 12246. .12334
/note="Similar to unknown repetitive sequence"

repeat_region complement(13006. .13517)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region complement(13638. .13727)
/note="Similar to unknown transposon"

repeat_region 13781. .13809
/note="Low_complexity AT_rich"

repeat_region 14700. .14759
/note="Simple_repeat (CGGGGG)n"

```
     gene            complement(15210. .17749)
                     /gene="OSJNBb0075K12.3"
                     /note="Hypothetical protein"
     CDS             complement(join(15210. .15407,17193. .17329,17428. .17749))
                     /gene="OSJNBb0075K12.3"
                     /codon_start=1
                     /product="Hypothetical protein"
                     /protein_id="AAM08569.1"
                     /db_xref="GI:19920137"
                     /translation="MGNCQAADAAAVVIQHPSSSSSSSSSSGNGGGGGGGRVERAYGA
                     VSAAAVMAANPGHYVAEVVRPVATAPAATAATASAPAARRRLKLLRPDDTLVLGGVYR
                     LVTFEDVLKQFVSKRNATMSRATIATAAEDDDGHRRQGHRGGEAAAAAPAKVAAQSRQ
                     ENPSSPSPTDEARPEPEPEPEPDLVATAMALGGRMSLSRHGQWRPALPSIAEGSVLCF
                     "
     repeat_region   15903. .16145
                     /note="Similar to putative MITE, MITE-adh, type B-like"
     repeat_region   16234. .16431
                     /note="Similar to putative MITE, Wanderer-like"
     repeat_region   complement(16500. .16527)
                     /note="Low_complexity AT_rich"
     repeat_region   17140. .17161
                     /note="Simple_repeat (CATCT)n"
     repeat_region   complement(17494. .17545)
                     /note="Simple_repeat (CGG)n"
     misc_feature    17625. .17680
                     /note="We believe the assembly to be correct.  The
                     assembly is covered by a single subclone on the top strand
                     with weak coverage by an additional subclone of less than
                     phred30 quality on the bottom strand.  The area is being
                     confirmed by PCR off of the BAC DNA."
     repeat_region   17670. .17700
                     /note="Simple_repeat (CGA)n"
     repeat_region   complement(18420. .18534)
                     /note="Similar to putative MITE, MITE-adh-1-like"
     gene            complement(19588. .23866)
                     /gene="OSJNBb0075K12.4"

  Query Match             91.7%;  Score 385;  DB 8;  Length 134933;
  Best Local Similarity   95.2%;  Pred. No. 6.4e-98;
  Matches  397;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  127196 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 127255

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  127256 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 127315

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  127316 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 127375

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db  127376 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 127435

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  127436 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 127495

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          |||||||||||| ||||||     |||||||||||| ||||||||||||||||||||||
Db  127496 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 127555

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||||| ||||||||||||||||||||
Db  127556 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 127612


RESULT 12
AP003280
LOCUS       AP003280               147201 bp    DNA     linear   PLN 17-APR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
            PAC clone:P0557A01.
ACCESSION   AP003280
VERSION     AP003280.2  GI:16554287
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:P0557A01.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0557A01
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 147201)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Oct 31, 2001 this sequence version replaced gi:13027310.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
             Detailed information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES                Location/Qualifiers
     source             1. .147201
                        /organism="Oryza sativa (japonica cultivar-group)"
                        /cultivar="Nipponbare"
                        /db_xref="taxon:39947"
                        /chromosome="1"
                        /clone="P0557A01"
     gene               join(4608. .4682,4991. .5422)
                        /gene="P0557A01.1"
     CDS                join(4608. .4682,4991. .5422)
                        /gene="P0557A01.1"
                        /note="hypothetical protein"
                        /codon_start=1
                        /protein_id="BAB89763.1"
                        /db_xref="GI:20160823"
                        /translation="MERGVAAARVGRRSRAGERRPIADGESFFSQQGLVSTHGYNMYT
                        RGHCKRSHRSIQLLSAHHHSLYRGFYLSFGGTWHLTRGCIGFRSPAKGSALRSAGPGN
                        CIGYVDIVQGCIVSTSWIALVRLIYLPTNLSYIYVNLVLAYRSTSIGCFSLYGICRYR
                        LNRLARLD"
     gene               join(7099.. .7303,8026. .8447)
                        /gene="P0557A01.2"
     CDS                join(7099. .7303,8026. .8447)
                        /gene="P0557A01.2"
                        /note="contains EST C74750(E50835)
                        similar to Arabidopsis thaliana chromosome 3, At3g20570
                        unknown protein"
                        /codon_start=1
                        /protein_id="BAB89764.1"
```

```
                     /db_xref="GI:20160824"
                     /translation="MAHGRVQRMQCWVVVACVVASVSTTASAFVFKAGGTGEWRVPDQ
                     QASGNVSAYNQWAEHTRFRVGDAIAFSYQPGNDSVLLVDKSSYDACNTNTPIDTFADG
                     NTVFTFTRSGPYYFISGNKDNCNRNEKLIVVVMGERAANGTAPAPALAPSAGTTSPNS
                     PPSPPPPPSGIEISPTPEQSINAAARPRAAGIAGAAGLAIGTLFYALV"
gene                 complement(10456. .10749)
                     /gene="P0557A01.3"
CDS                  complement(10456. .10749)
                     /gene="P0557A01.3"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB89765.1"
                     /db_xref="GI:20160825"
                     /translation="MHASKSSRFSSVYAKIMDQGNMTLTKKKLIYRTEVCPYVLTFCE
                     KSFFYPNSFGTQAQNSQEKNYEILGENTGSVAVPSPAKPFIKATKDSLPSLCC"
gene                 join(11771. .11804,12284. .12312,12635. .12792,13516. .13633,
                     13717. .13905,14205. .14540,14665. .14766,14893. .15087,
                     15190. .15352,15511. .15776,15867. .15962,16057. .16473)
                     /gene="P0557A01.4"
CDS                  join(11771. .11804,12284. .12312,12635. .12792,13516. .13633,
                     13717. .13905,14205. .14540,14665. .14766,14893. .15087,
                     15190. .15352,15511. .15776,15867. .15962,16057. .16473)
                     /gene="P0557A01.4"
                     /note="contains ESTs C74009(E30143),AU029314(E30143)"
                     /codon_start=1
                     /product="putative homeobox 1"
                     /protein_id="BAB89766.1"
                     /db_xref="GI:20160826"
                     /translation="MGDEWETWIAADMGQNKGALINDGEQRMSIPARELFAAIEADSG
                     LLPVNSSNTNEKRKRRLQRLTGKQSEVLEGFFSICGHPDDGQKRHLSETTGLGLDQVK
                     FWFQNKRTQVKTMCWKEENYKLSVENEILRDENRRVKIAHCTAVCLTCCNSSVQNQLA
                     VEMERLMGQSEWLQQEIARSNGTPPAANLAFQLNSSADYVFSGQHDQQMIAELAKNAM
                     HALIILAESHVALWFPVPGCSYEVLNKMAYDQAYPGDNSANAIGFKTEATRAVSMVMM
                     DYKSVVDFLMDPYNYRTFFPEVISGAVTNRIYTWPTSDGYNGVIQLMTVEMMFPSPLV
                     PARKCTFLRYCNVLNEGLVVVIDVSLDDGSIFSKCRKMPSGFLIQSIRPNSCKVTAIE
                     HVLADDTGVHELYQPCMNGLVFGARRWVATMARQSARMRDVHHNKTAPQVSTKGRKNL
                     MKLADDLLASFAGGIAATGGGTWTVVIGAGTEKDIRVAYRRTTEGSSSYNAILSVTAS
                     LRLPLPMRKTFDLLRNLTHRCKWDVLVHGSVVKEEVTIARGVGNDDTVTVLHCKRAGR
                     EDRGRTMILQNNGYDASGSFMVYSQIDSELMNTMVLSPSDLPPGRGGPSLYPTGFSLL
                     PDVEAAQDSSGIALGEVGGTLMTMGFQIPVKLASGDRMYSRSAASAIRLMTDTIALVK
                     KTLMNEHSGIYGVSPFHP"
gene                 complement(17785. .19797)
                     /gene="P0557A01.5"
CDS                  complement(17785. .19797)
                     /gene="P0557A01.5"
                     /note="hypothetical protein
                     similar to Arabidopsis thaliana chromosome 1, F23O10.12"
                     /codon_start=1
                     /protein_id="BAB89767.1"
                     /db_xref="GI:20160827"
                     /translation="MRALLRLRRRLPLPINTRAFSPSSPSPAPHEIPTVYSFLQPSVF
                     APRPKPQPPPPPPPPTPPAHKTLPVGDAVALEDELLAAVSEDRSDDAWLAFRSLASAS
                     LSPSPPAAAALVSHLAAAHHHRLGLKRAFAAAVFLLEKSPHADPVPEAALQAVFTSLA
                     AAASAAPALALVRALLRCGRRLPAFPVWGSPLIELTRADTGAFVAFLKVFDEACKQMV
                     SEKSPSAAAAMRPDLAACNAVLGGCCRLLGSVTEAERVLEIMSAIAVSPDVDSFGCLA
                     FLYAWRDIPSRVDELDKLLDALGFGKKIFFKNLISGYLKSCSFESVSSVILRVVEERR
                     VGGSNAFDLESYTEVAQRFVDNGRIRELAQLIIKAQETESLQQSLAVEDSVGFGIVNA
                     CVELGLLNKAHSILDEMTAQGASVGLGVYSSILKAYCKEQRTAEAAQLVSEISAAGLQ
                     LDAGSYDALIDASMTAHDFLSAFSLFKEMREARLPDLRTSYLTIMTGLTENNRPELMA
                     SFLDTVVDDPRIEIATHDWNSIIHAFCKVGRLEDARRTYRRMVFLRYEPNNQTYLSLI
                     NGYVSAEKYFSVLILWTEVRRKGADFNHELIDAFLYALVKGGFFDMAMQVIEKAQELK
                     IFLDKWRHKQAFMETHKKLKVAKLRKRNFRKMEALIAFKNWAGLNA"
gene                 complement(join(21056. .21478,22405. .22644))
CDS                  complement(join(21056. .21478,22405. .22644))
                     /gene="P0557A01.6"
                     /note="contains ESTs C98021(C0384),D15269(C0384)
                     similar to Arabidopsis thaliana chromosome 3, At3g13230
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAB89768.1"
                     /db_xref="GI:20160828"
                     /translation="MEVDRAASAAAAGESSSAGAMAVEKPRFEALMPSEMSGGRPQFR
                     KVPVPQHRFAPLKKAWMDIYTPVYEHMKIDIRMNLKARRVELKTRQDTPDVSNLQKCA
                     DFVHAFMLGFDIADAVALLRLDDLYVDSFEIKDVKTLRGEHLSRAIGRLSGKGGKTKY
                     AIENSTRTRIVIADTKIHILGSFVNIKVARDSLCSLILGSPAGKVYSKLRAVSARLAE
                     RY"
gene                 complement(join(28212. .28268,28479. .28512,28787. .28881,
                     28966. .29040,29122. .29207,29453. .29504,29578. .29652,
                     29751. .29817,30896. .31356))
                     /gene="P0557A01.7"
CDS                  complement(join(28212. .28268,28479. .28512,28787. .28881,
                     28966. .29040,29122. .29207,29453. .29504,29578. .29652,
                     29751. .29817,30896. .31356))
                     /gene="P0557A01.7"
                     /note="hypothetical protein
                     similar to Arabidopsis thaliana chromosome 4, At4g04870"
                     /codon_start=1
                     /protein_id="BAB89769.1"
                     /db_xref="GI:20160829"
                     /translation="MAFLKTLNPLLRRSPTPIPNPRSLLSLDAFLAASSPTAASHATA
                     PAPFAAAAHHHVPIRSGGPLFLSSPPWMLSQSATPLTAAAAALRARLRRARALAGGGA
                     QAVADAVGWEPRRISRDESEVAEAVTGGRERFLNLPNLVSIGRMASGPVIGWMIVNEW
                     YLPAFGTLALSGASDWLDGFLARKMGINSVFGSYLDPLADKVLIGCVAIAMVEKDLLH
                     PGLVGLVVVRDLLLVGGAVYKRASSLGWKWNSWSDFVNLDAIHREKVKPLFISKVNTV
                     FQLMLVAAALLQPEFGTEETQNYITVLSWLVASTTIASTLFVKHLEQDTEGRGQTSE"
gene                 join(33110. .33437,33549. .33727,33814. .34003,34085. .34476,
                     34564. .34878)
                     /gene="P0557A01.8"
CDS                  join(33110. .33437,33549. .33727,33814. .34003,34085. .34476,
                     34564. .34878)
                     /gene="P0557A01.8"
                     /codon_start=1
                     /product="protein kinase-like"
                     /protein_id="BAB89770.1"
                     /db_xref="GI:20160830"
                     /translation="MDLLTSAMKRCCNGKTLLRGGRTCRCSSASAIGSSGVRGKEEAS
                     TSAADSEPDKKRWRKKRFWRKKKKKAKKDHYGDAAATEHGSERASCRRYENDAVADLV
                     NDISSKSDVCNVYAAEGILRITHQNIPSMVLTYRQLCNATDSFSPNNLLGEGGFGRVY
                     RGHLEEINEIVAVKQLDKDGFQGNREFLVEVLMLSLLHHPNLVKLLGYCTDMDQRILV
                     YECMRNGSLEDHLLDLPPKAKPLPWQTRMKIAVGAAKGIEYLHEVANPPVIYRDLKTS
                     NILLDEDFNSKLSDFGLAKLGPVGDKSHVSTRVMGTYGYCAPEYAMTGKLTKTSDIYS
                     FGVVLLEIITGRRAIDTSRPTHEQVLVQWAAPLVKDKKRFVRLADPLLEEKFPLKGLY
                     QALAIASMCLQEDASNRPMISDVVAALSFLAEQKYHPQDGPDQAARKSRDRDCSNPPR
                     KTDMVSEIKADDEIKHR"
gene                 36968. .37666
                     /gene="P0557A01.9"
CDS                  36968. .37666
                     /gene="P0557A01.9"
                     /codon_start=1
                     /product="apocytochrome b6"
                     /protein_id="BAB89771.1"
                     /db_xref="GI:20160831"
                     /translation="MKFSYTVLGGGFGLVTYLNKVYDWFEERLEIQAIADDITSKYVP
                     PHVNIFYCLGGITLTCFLVQVATGFAMTFYYRPTVTEAFSSVQYIMTEANFGWLIRSV
                     HRWSASMMVLMMILHVFRVYLTGGFKKPRELTWVTGVVLAVLTASFGVTGYSLPWDQI
```

Query Match 91.7%; Score 385; DB 8; Length 147201;
Best Local Similarity 95.2%; Pred. No. 6.4e-98;
Matches 397; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   92584 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 92643

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   92644 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 92703

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   92704 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 92763

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||||
Db   92764 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 92823
```

```
Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   92824 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 92883

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||  ||||||     ||||||||||  ||||||||||||||||||||||
Db   92884 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 92943

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           ||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db   92944 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 93000


RESULT 13
AP003986
LOCUS       AP003986              105649 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 6 clone
            OJ1057_A09, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP003986
VERSION     AP003986.1  GI:15076789
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OJ1057_A09.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, BAC
            clone:OJ1057_A09
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 105649)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .105649
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="6"
                     /clone="OJ1057_A09"
BASE COUNT    32062 a  21297 c  21329 g  30937 t     24 others
ORIGIN

  Query Match             91.3%;  Score 383.4;  DB 2;  Length 105649;
  Best Local Similarity   95.0%;  Pred. No. 1.8e-97;
  Matches  396;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   13085 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 13144

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13145 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 13204

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   13205 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 13264

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db   13265 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 13324

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Db   13325 ACAACTGTTTGAACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 13384

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||| ||||||     |||||||||| ||||||||||||||||||||||||
Db   13385 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 13444

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           ||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db   13445 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 13501


RESULT 14
AP003623/c
LOCUS       AP003623              139914 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 6 clone P0642B07,
            *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP003623
VERSION     AP003623.1  GI:14020961
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:P0642B07.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
            clone:P0642B07
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 139914)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .139914
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="6"
                     /clone="P0642B07"
BASE COUNT    41395 a  28837 c  28251 g  41381 t     50 others
ORIGIN

  Query Match             91.3%;  Score 383.4;  DB 2;  Length 139914;
  Best Local Similarity   95.0%;  Pred. No. 1.8e-97;
  Matches  396;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;
```

```
Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   39796 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 39737

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39736 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 39677

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   39676 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 39617

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db   39616 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 39557

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Db   39556 ACAACTGTTTGAACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 39497

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           |||||||||||| ||||||     |||||||||||  |||||||||||||||||||||||
Db   39496 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 39437

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  ||||||||||||||||| ||||||||||||||||||||
Db   39436 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 39380
```

RESULT 15
AP004236/c
LOCUS AP004236 167379 bp DNA linear HTG 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone P0417D05, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AP004236
VERSION AP004236.1 GI:15919947
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:P0417D05.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC clone:P0417D05
JOURNAL Published Only in Database (2001)
REFERENCE 2 (bases 1 to 167379)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES Location/Qualifiers
source 1. .167379
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="6"
/clone="P0417D05"
BASE COUNT 48364 a 34609 c 34564 g 49521 t 321 others
ORIGIN

Query Match 91.3%; Score 383.4; DB 2; Length 167379;
Best Local Similarity 95.0%; Pred. No. 1.8e-97;
Matches 396; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 157309 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 157250

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157249 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 157190

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db 157189 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 157130

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           |||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db 157129 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 157070

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Db 157069 ACAACTGTTTGAACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 157010

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           |||||||||||| ||||||     |||||||||||  |||||||||||||||||||||||
Db 157009 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 156950

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  ||||||||||||||||| ||||||||||||||||||||
Db 156949 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 156893
```

Search completed: April 15, 2003, 18:31:33
Job time : 23702 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 12:46:46 ; Search time 3925 Seconds
(without alignments)
240.978 Million cell updates/sec

Title: US-09-696-664A-3
Perfect score: 420
Sequence: 1 gcgcgttgtgagaattctta.....tcattgtgggtaacgtattt 420

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

[redacted]tched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	385	91.7	134525	11	AAQ04525	Total base sequenc
	2	340.8	81.1	4260	22	AAF84278	rbcL-ORF512 sequen
	3	177	42.1	1416	16	AAQ87822	Agmenellum quadrup
	4	177	42.1	4234	16	AAQ87819	Agmenellum quadrup
	5	133.2	31.7	2222	13	AAQ22147	RuBPCase gene. Th
	6	129.4	30.8	1419	24	ABQ90432	M. capsulatus gene
	7	80.4	19.1	1953	21	AAZ61381	Nucleotide sequenc
	8	80	19.0	195	21	AAZ61365	Nucleotide sequenc
	9	64	15.2	195	21	AAZ61367	Nucleotide sequenc
	10	44.2	10.5	159	21	AAZ61366	Nucleotide sequenc
	11	41	9.8	1416	20	AAX21432	Regulatory region
	12	41	9.8	1417	18	AAT85196	Plastid targeting
	13	39.2	9.3	278	20	AAX21420	5' and 3' regulato
	14	38	9.0	6433	20	AAX13041	Enterococcus faeca
c	15	36.6	8.7	349980	22	AAH41225	Pyrococcus abyssi
	16	36.2	8.6	191	20	AAX21410	PtrnV promoter seq
	17	35.6	8.5	2543	21	AAA52774	Soybean putative c
c	18	35.6	8.5	3746	21	AAZ99512	DNA encoding a mai
c	19	35.6	8.5	3746	21	AAZ99527	DNA encoding a mai
c	20	35.6	8.5	3773	21	AAZ99494	DNA encoding a mai
c	21	35	8.3	4270	24	ABI99885	Mouse ischaemic co
	22	34.8	8.3	22008	22	ABA15839	Human nervous syst
c	23	34.8	8.3	22008	22	AAS36610	Human cardiovascul
	24	34.4	8.2	165	20	AAX21409	Prrn promoter sequ
	25	34.4	8.2	465237	24	ABQ87681	Human oestrogen re
	26	34.4	8.2	465237	24	ABA90193	Human oestrogen re
	27	34.2	8.1	3381	19	AAV41343	M. catarrhalis str
c	28	33.4	8.0	337	22	AAK67586	Human immune/haema
c	29	33.4	8.0	473	22	AAK56227	Human immune/haema
c	30	33.4	8.0	3595	23	ABL12127	Drosophila melanog
	31	33	7.9	1827	24	ABQ70608	Listeria monocytog
c	32	32.8	7.8	495	23	ABK42341	Genomic sequence #
c	33	32.8	7.8	495	23	ABK42342	Genomic sequence #
	34	32.6	7.8	2373	22	AAH65995	C glutamicum codin
c	35	32.6	7.8	349980	22	AAH68527	C glutamicum codin
	36	32.4	7.7	568	21	AAA44116	Human secreted exp
	37	32.4	7.7	3191	21	AAZ58978	Human cytoskeletal
	38	32.4	7.7	3363	22	AAS45004	cDNA encoding nove
	39	32.2	7.7	580073	18	AAT58840	Mycoplasma genital
c	40	32	7.6	1693	21	AAC59175	Human secreted pro
c	41	32	7.6	1696	24	ABQ54262	Human ovarian anti
c	42	32	7.6	2807	24	ABK48311	DNA encoding cap-b
c	43	32	7.6	2992	24	ABN85281	Cell cycle regulat
c	44	32	7.6	5954	19	AAV59105	Zebrafish differen
	45	31.8	7.6	168	20	AAX21407	Prrn/rbcL/Rubisco

ALIGNMENTS

RESULT 1
AAQ04525
ID AAQ04525 standard; DNA; 134525 BP.
XX
AC AAQ04525;
XX
DT 01-OCT-1990 (first entry)
XX
DE Total base sequence of rice plant chloroplast DNA.
XX
KW Chloroplast; true grass; rice plant; ss.
XX
OS Oryza sativa.
XX
PN JP02100682-A.
XX
PD 12-APR-1990.
XX
PF 07-OCT-1988; 88JP-0251967.
XX
PR 07-OCT-1988; 88JP-0251967.
XX
PA (MITK) MITSUI TOATSU CHEM INC.
XX
DR WPI; 1990-159709/21.
XX
PT Chloroplast DNA of true grasses - used to produce various
PT DNA base sequences by decomposition of rice plant DNA.
XX
PS Claim 1; Fig 1; 20pp; Japanese.
XX
CC The sequence is that of the whole of rice chloroplast DNA.

```
XX
SQ   Sequence 134525 BP; 41249 A; 26129 C; 26331 G; 40816 T; 0 other;

  Query Match             91.7%;  Score 385;  DB 11;  Length 134525;
  Best Local Similarity   95.2%;  Pred. No. 2.6e-107;
  Matches  397;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   54053 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACGTATGTCACCACAAACAGAA 54112

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54113 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 54172

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   54173 CCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG 54232

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||||||||| |||||||||| |||||||||||||| ||||||||||||||||||||||
Db   54233 GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG 54292

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   54293 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCAC 54352

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||  ||||||     ||||||||||  ||||||||||||||||||||||
Db   54353 ATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGAC 54412

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  ||||||||||||||| ||||||||||||||||||||
Db   54413 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 54469


RESULT 2
AAF84278
ID   AAF84278 standard; DNA; 4260 BP.
XX
AC   AAF84278;
XX
DT   13-JUN-2001  (first entry)
XX
DE   rbcL-ORF512 sequence.
XX
KW   Plant; polyester synthase; alkyl 3-hydroxyalkanoic acid;
KW   rbcL-ORF512; plastid; ds.
XX
OS   Nicotiana tabacum.
XX
[redacted]   EP1076095-A1.
XX
PD   14-FEB-2001.
XX
PF   08-AUG-2000; 2000EP-0117037.
XX
PR   09-AUG-1999;   99JP-0225832.
PR   09-AUG-1999;   99JP-0225839.
XX
PA   (RIKE ) RIKEN KK.
XX
PI   Yamaguchi I,  Nakashita H,  Yoshioka K,  Doi Y;
XX
DR   WPI; 2001-292601/31.
XX
PT   Transforming plants for producing polyester, involves ligating an
PT   operon containing a promoter and several genes of interest to a vector
PT   and integrating the recombinant vector into a plastid chromosome  -
XX
PS   Example 1; Page 26-28; 37pp; English.
XX
CC   The present invention relates to methods for transformation of plants.
CC   The methods comprise ligating an operon containing a promoter and 2-100
CC   genes of interest, to a vector and integrating the resulting recombinant
CC   vector into a plastid chromosome. The methods are useful for transforming
CC   a plant, belonging to any family such as Gramineae, Malvaceae,
CC   Brassicaceae, Compositae, Pedaliaceae, Oleaceae, Myrtaceae, Rosaceae,
CC   Theaceae, Leguminosae, Palmae, Sterculiaceae or Rubiaceae, in particular
CC   Nicotiana tabacum of Solanaceae family, for producing polyester, which is
CC   copolymer of lower alkyl 3-hydroxyalkanoic acid. In the present invention
CC   polyester synthase genes (see AAF84261-AAF84264) were used in the method
CC   to generate a plant which produces polyester. In order to ensure that
CC   homologous recombinantion is carried out in the plastid chromosome, the
CC   present sequence, the rbcL and ORF512 genes, which exist in the plastid
CC   were used.
XX
SQ   Sequence 4260 BP; 1336 A; 675 C; 892 G; 1357 T; 0 other;

  Query Match             81.1%;  Score 340.8;  DB 22;  Length 4260;
  Best Local Similarity   88.7%;  Pred. No. 2.4e-94;
  Matches  369;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAA 64
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||| |
Db     186 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGA 245

Qy      65 CTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCC 124
          |||||||||||||||||| |||||||||||||| || || |||||||||| || || |
Db     246 CTAAAGCAAGTGTTGGATTCAAAGCTGGTGTTAAAGAGTACAAATTGACTTATTATACTC 305

Qy     125 CGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCG 184
          | |||||| |||||||||||||||||| |||||||||||||||||||||||||||| | |
Db     306 CTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCCGAGTAACTCCTCAACCTG 365

Qy     185 GGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGA 244
          | ||||| |||||||||||||| || || |||||||| ||||||||||||||||||||||
Db     366 GAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGA 425

Qy     245 CAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACA 304
          ||||||| ||||| |||||||||||||| |||||||||||||||| |||||||| | ||
Db     426 CAACTGTATGGACCGATGGACTTACCAGCCTTGATCGTTACAAAGGGCGATGCTACCGCA 485

Qy     305 TCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACC 364
          ||||||  |||  ||| ||   |||||||||||   |||||||||||| || |||||||
Db     486 TCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGCTTACCCTTTAGACC 545

Qy     365 TATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          | |||||||| |||||  ||| |||||||||||| |||||| ||||||||||||
Db     546 TTTTTGAAGAAGGTTCTGTTACCAACATGTTTACTTCCATTGTAGGTAACGTATTT 601


RESULT 3
AAQ87822
ID   AAQ87822 standard; DNA; 1416 BP.
XX
AC   AAQ87822;
XX
DT   14-DEC-1995  (first entry)
XX
DE   Agmenellum quadruplicatum PR-6 rubisco large subunit coding region.
XX
KW   Rubisco; ribulose 1,5-bisphosphate carboxylase/oxygenase;
KW   large subunit; carbon dioxide fixation; Synechococcus; ds.
XX
OS   Agmenellum quadruplicatum PR-6.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1416
FT                   /*tag=  a
FT                   /product= large_subunit
XX
PN   JP07079782-A.
XX
PD   28-MAR-1995.
XX
```

```
PF   18-JUN-1993;   93JP-0184304.
XX
PR   18-JUN-1993;   93JP-0184304.
XX
PA   (KANT ) KANSAI DENRYOKU KK.
XX
DR   WPI; 1995-157852/21.
DR   P-PSDB; AAR72392.
XX
PT   Promoter sequence of Synechococcus PCC7002-originated rubisco gene -
PT   used in a vector to produce Cyano:bacterium with improved carbon
PT   di:oxide fixation
XX
PS   Claim 7; Page 17-18; 28pp; Japanese.
XX
CC   A sequence including the ORFs coding for the large and small
CC   subunits of the rubisco enzyme has been isolated from Agmenellum
CC   quadruplicatum PR-6 (Synechococcus sp. strain PCC7002). The
CC   region coding for the rubisco large subunit is claimed; the enzyme
CC   is involved in carbon dioxide fixation.
XX
SQ   Sequence 1416 BP; 303 A; 405 C; 346 G; 362 T; 0 other;

  Query Match           42.1%;  Score 177;  DB 16;  Length 1416;
  Best Local Similarity 69.6%;  Pred. No. 3.9e-44;
  Matches  240;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy      75 TGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGA 134
           || ||| ||||| || |||||  |||| ||     ||||||||||||||| || |||
Db      18 TGCTGGGTTTAATGCCGGTGTACAGGACTACCGCCTGACTTACTACACCCCCGATTACAC 77

Qy     135 AACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGGGGTTCCGCC 194
             | || ||||| ||  |  | ||    |||||  | ||||| || |  || || || ||
Db      78 CCCGAAAGATACCGACTTACTCGCTTGTTTCCGGATGACTCCCCAACCTGGAGTCCCCCC 137

Qy     195 TGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGACAACTGTTTG 254
           ||||||   |  || || || || || |||||||||||| ||||| ||||| ||||| ||
Db     138 CGAAGAATGTGCTGCGGCTGTTGCGGCTGAATCTTCTACCGGTACTTGGACCACTGTATG 197

Qy     255 GACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACATCGAGCCCGT 314
           ||| |||||  | ||    || || || ||||| || || |||||  |  | || |||||
Db     198 GACCGATGGTTTAACTGACCTCGACCGCTACAAGGGTCGTTGCTACAATGTTGAACCCGT 257

Qy     315 TCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCTATTTGAAGA 374
           ||| || ||     | |||||| ||||||  || ||||| ||  | || || ||||||||
Db     258 TCCCGGTGAAGACAACCAATATTTCTGTTTCGTTGCTTACCCCCTCGATCTGTTTGAAGA 317

Qy     375 GGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATT 419
           ||||||   || ||| | || |||| | | || ||||||||||
Db     318 AGGTTCTGTAACCAACGTTTTGACTTCCTTGGTTGGTAACGTATT 362


RESULT 4
AAQ87819
ID   AAQ87819 standard; DNA; 4234 BP.
XX
AC   AAQ87819;
XX
DT   14-DEC-1995  (first entry)
XX
DE   Agmenellum quadruplicatum PR-6 rubisco gene.
XX
KW   Rubisco; ribulose 1,5-bisphosphate carboxylase/oxygenase; promoter;
KW   carbon dioxide fixation; Synechococcus sp.; ds.
XX
OS   Agmenellum quadruplicatum PR-6.
XX
FH   Key             Location/Qualifiers
FT   CDS             45..344
FT                   /*tag=  a
FT                   /label= ORF3
FT                   /note= "claim 5, encodes AAR72391"
FT   promoter        659..962
FT                   /*tag=  b
FT                   /note= "claim 1"
FT   CDS             963..2378
FT                   /*tag=  c
FT                   /product= large_subunit
FT                   /note= "claim 7, encodes AAR72392"
FT   CDS             2447..2851
FT                   /*tag=  d
FT                   /label= ORF2
FT                   /note= "claim 9, encodes AAR72393"
FT   CDS             2882..3217
FT                   /*tag=  e
FT                   /product= small_subunit
FT                   /note= "claim 11, encodes AAR72394"
FT   CDS             3379..4005
FT                   /*tag=  f
FT                   /label= ORF1
FT                   /note= "claim 13, encodes AAR72395"
FT
XX
PN   JP07079782-A.
XX
PD   28-MAR-1995.
XX
PF   18-JUN-1993;   93JP-0184304.
XX
PR   18-JUN-1993;   93JP-0184304.
XX
PA   (KANT ) KANSAI DENRYOKU KK.
XX
DR   WPI; 1995-157852/21.
DR   P-PSDB; AAR72391, AAR72392, AAR72393, AAR72394, AAR72395.
XX
PT   Promoter sequence of Synechococcus PCC7002-originated rubisco gene -
PT   used in a vector to produce Cyano:bacterium with improved carbon
PT   di:oxide fixation
XX
PS   Example 7; Page 9-14; 28pp; Japanese.
XX
CC   A sequence including the ORFs coding for the large and small
CC   subunits of the rubisco enzyme has been isolated from Agmenellum
CC   quadruplicatum PR-6 (Synechococcus sp. strain PCC7002). The
CC   promoter region and various coding regions are each claimed (see
CC   Features Table). The rubisco enzyme is involved in carbon dioxide
CC   fixation.
XX
SQ   Sequence 4234 BP; 1142 A; 1118 C; 900 G; 1074 T; 0 other;

  Query Match           42.1%;  Score 177;  DB 16;  Length 4234;
  Best Local Similarity 69.6%;  Pred. No. 6.2e-44;
  Matches  240;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy      75 TGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGA 134
           || ||| ||||| || |||||  |||| ||     ||||||||||||||| || |||
Db     980 TGCTGGGTTTAATGCCGGTGTACAGGACTACCGCCTGACTTACTACACCCCCGATTACAC 1039

Qy     135 AACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGGGGTTCCGCC 194
             | || ||||| ||  |  | ||    |||||  | ||||| || |  || || || ||
Db    1040 CCCGAAAGATACCGACTTACTCGCTTGTTTCCGGATGACTCCCCAACCTGGAGTCCCCCC 1099

Qy     195 TGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGACAACTGTTTG 254
           ||||||   |  || || || || || |||||||||||| ||||| ||||| ||||| ||
Db    1100 CGAAGAATGTGCTGCGGCTGTTGCGGCTGAATCTTCTACCGGTACTTGGACCACTGTATG 1159

Qy     255 GACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACATCGAGCCCGT 314
           ||| |||||  | ||    || || || ||||| || || |||||  |  | || |||||
Db    1160 GACCGATGGTTTAACTGACCTCGACCGCTACAAGGGTCGTTGCTACAATGTTGAACCCGT 1219

Qy     315 TCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCTATTTGAAGA 374
           ||| || ||     | |||||| ||||||  || ||||| ||  | || || ||||||||
Db    1220 TCCCGGTGAAGACAACCAATATTTCTGTTTCGTTGCTTACCCCCTCGATCTGTTTGAAGA 1279

Qy     375 GGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATT 419
           ||||||   || ||| | || |||| | | || ||||||||||
```

```
Db   1280 AGGTTCTGTAACCAACGTTTTGACTTCCTTGGTTGGTAACGTATT 1324
```

RESULT 5
AAQ22147
ID AAQ22147 standard; DNA; 2222 BP.
XX
AC AAQ22147;
XX
DT 30-JUN-1992 (first entry)
XX
DE RuBPCase gene.
XX
KW D-ribulose-1,5-bisphosphate carboxylase; Rubisco; bioreactor; carbon;
KW dioxide absorption; large subunit; small subunit; ss.
XX
OS Thiobacillus ferrooxidans.
XX

FH	Key	Location/Qualifiers
FT	CDS	181..1603
FT		/*tag= a
FT		/product= large_subunit
FT	CDS	1693..2026
FT		/*tag= b
FT		/product= small_subunit

XX
PN JP04030792-A.
XX
PD 03-FEB-1992.
XX
PF 29-MAY-1990; 90JP-0139300.
XX
PR 29-MAY-1990; 90JP-0139300.
XX
PA (DOWA) DOWA MINING CO LTD.
PA (AKIT) AKITA KEN.
XX
DR WPI; 1992-085966/11.
DR P-PSDB; AAR21977; AAR23903.
XX
PT RuBPCase gene from Thiobacillus ferrooxidans - used to produce
PT RuBPCase for CO2 absorption bioreactor
XX
PS Claim 1; Page 1; 16pp; Japanese.
XX
CC The D-ribulose-1,5-bisphosphate carboxylase (RuBPCase) gene was
CC isolated from Thiobacillus ferrooxidans. The genomic DNA of T.
CC ferrooxidans was recovered from a 9K medium culture incubated with
CC Proteinase K. A RuBPCase gene fragment from Chromatium vinosum was
CC used as a probe in Southern hybridisation to genomic DNA. The T.
CC ferrooxidans Fel strain RuBPCase gene was cloned and a deletion
CC plasmid constructed. The nucleotide and corresp. amino acid
CC sequences of RuBPCase were determined. The gene encodes both the
CC large and small RuBPCase subunits, which together form a complex of
CC 8 large and 8 small subunits designated Rubisco (EC.4.1.1.39).
CC An E.coli plasmid for expressing the RuBPCase gene was constructed
CC and the RuBPCase activity in E. coli determined. RuBPCase produced
CC by E. coli or Thiobacillus can be used in a bioreactor of carbon
CC dioxide absorption.
XX
SQ Sequence 2222 BP; 496 A; 623 C; 657 G; 446 T; 0 other;

Query Match 31.7%; Score 133.2; DB 13; Length 2222;
Best Local Similarity 62.1%; Pred. No. 1.4e-30;
Matches 210; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy     82 TTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAAACCAAG 141
Db    196 TATGAGGCCGGTGTAAAGGACTACCGGCAGACCTATTGGGCCCCCGAGTACGTGCCTTTG 255

Qy    142 GATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGGGGTTCCGCCTGAAGAA 201
Db    256 GATTCCGATATCCTAGCCTGTTTCAAGATTACGCCGCAACCAGGGGTGGATCGTGAGGAA 315

Qy    202 GCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGACAACTGTTTGGACTGAT 261
Db    316 GCCGCCGCCGCGGTAGCCGCCGAGTCTTCTACCGGCACATGGACCACGGTGTGGACCGAT 375

Qy    262 GGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACATCGAGCCCGTTCCTGGG 321
Db    376 TTGTTGACCGACATGGATTACTACAAAGGCCGCGCCTATCGCATTGAAGATGTGCCGGGC 435

Qy    322 GACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCTATTTGAAGAGGGTTCT 381
Db    436 GACGACACCTGCTTTTACGCGTTCGTTGCCTATCCCATAGATTTGTTTGAAGAGGGGTCC 495

Qy    382 TGTACTAACATGTTTACTTTCATTGTGGGTAACGTATT 419
Db    496 GTGGTTAACGTCTTCACCTCCCTGGTCGGTAACGTGTT 533
```

RESULT 6
ABQ90432
ID ABQ90432 standard; DNA; 1419 BP.
XX
AC ABQ90432;
XX
DT 01-OCT-2002 (first entry)
XX
DE M. capsulatus gene #417 for DNA array.
XX
KW Micro array; gene; ds; differential expression; gene expression.
XX
OS Methylococcus capsulatus.
XX
PN WO200255655-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002WO-NO00019.
XX
PR 12-JAN-2001; 2001NO-0000235.
PR 12-JAN-2001; 2001NO-0000239.
XX
PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
DR WPI; 2002-557818/59.
XX
PT Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes -
XX
PS Claim 13; Page 248; 678pp; English.
XX
CC The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
CC invention.
XX
SQ Sequence 1419 BP; 285 A; 493 C; 430 G; 211 T; 0 other;

Query Match 30.8%; Score 129.4; DB 24; Length 1419;
Best Local Similarity 61.4%; Pred. No. 1.7e-29;
Matches 208; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy     81 ATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAAACCAA 140
```

```
Db      15 ATACAACGCGGGCGTCAAGGAATACCGCGAAACCTACTGGGACCCGAACTACACTCCCGC 74

Qy     141 GGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGGGGTTCCGCCTGAAGA 200
           || || ||| |  |||| |  |||     | || |||||||  || || ||||  |||||
Db      75 CGACACCGATCTGCTGGCGGTCTTCAAGATCACCCCTCAGCCGGGTGTGCCGCGCGAAGA 134

Qy     201 AGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGACAACTGTTTGGACTGA 260
           |||  |  || || || || |||||||| || || || || ||||| || || ||||| ||
Db     135 AGCCGCCGCCGCCGTGGCCGCGGAATCGTCGACCGGCACCTGGACCACCGTCTGGACCGA 194

Qy     261 TGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACATCGAGCCCGTTCCTGG 320
             || |||   || ||   |||||| || ||   ||| |  ||||||  ||| || ||
Db     195 CCTGCTGACCGACCTCGACTATTACAAGGGCCGCGCCTACCGGATCGAGGACGTGCCCGG 254

Qy     321 GGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCTATTTGAAGAGGGTTC 380
             |   || || |    |   |   | || |||||  | ||||| || || ||||| ||
Db     255 CCAGGACGAACAGTTCTACGCCTTCATCGCCTATCCCATCGACCTGTTCGAGGAGGGCTC 314

[redacted] 381 TTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATT 419
               ||| |||| || |   | ||||| ||||| ||
Db     315 CGTCGTCAACGTGTTCACCTCGCTGGTGGGCAACGTGTT 353
```

```
RESULT 7
AAZ61381
ID   AAZ61381 standard; DNA; 1953 BP.
XX
AC   AAZ61381;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Nucleotide sequence of DNA construct FLARE16-S1.
XX
KW   Green fluorescent protein; GFP; aadA; aptB; protein expression; vaccine;
KW   haemoglobin; enzyme; psbA; ss.
XX
OS   Synthetic.
OS   Unidentified.
OS   Aequorea victoria.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    7..189
FT                   /*tag=  a
FT                   /note= "atpB downstream box"
FT   misc_feature    196..978
FT                   /*tag=  b
[redacted]           /note= "aadA sequence"
     misc_feature    1027..1747
FT                   /*tag=  c
FT                   /note= "green fluorescent protein region"
FT   misc_feature    1758..1946
FT                   /*tag=  d
FT                   /note= "psbA region"
XX
PN   WO200007431-A1.
XX
PD   17-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17806.
XX
PR   03-AUG-1998;   98US-0095163.
PR   03-AUG-1998;   98US-0095167.
PR   15-DEC-1998;   98US-0112257.
PR   29-APR-1999;   99US-0131611.
PR   11-JUN-1999;   99US-0138764.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Maliga P,  Kuroda H,  Khan MS;
XX
DR   WPI; 2000-205525/18.
XX
PT   New recombinant DNA constructs, for expressing high levels of
PT   heterologous protein in plastids of higher plants, includes promoter, a
PT   leader sequence and a downstream box element -
XX
PS   Disclosure; Fig 29; 164pp; English.
XX
CC   The present sequence represents a DNA construct of the invention.
CC   The specification describes recombinant DNA constructs for expressing
CC   heterologous proteins in the plastids of higher plants. The DNA
CC   constructs comprise a 5' regulatory region which includes a promoter
CC   element, a leader sequence and a downstream box element operably linked
CC   to a coding region of the heterologous protein. The chimeric regulatory
CC   region enhances translational efficiency of an mRNA molecule encoded by
CC   the DNA construct. The DNA constructs are used for producing transformed
CC   monocot and dicot plants having high levels of heterologous protein
CC   expression. They can be used to drive expression of proteins with
CC   agronomic, industrial or pharmaceutical importance, including production
CC   of vaccines, healthcare products like human haemoglobin, industrial or
CC   household enzymes. Plants which can be transformed with the constructs
CC   of the invention include maize, millet, sorghum, sugar cane, rice,
CC   wheat, barley, oat, rye or turf grass.
XX
SQ   Sequence 1953 BP; 546 A; 422 C; 487 G; 498 T; 0 other;

  Query Match             19.1%;  Score 80.4;  DB 21;  Length 1953;
  Best Local Similarity   93.3%;  Pred. No. 2.4e-14;
  Matches   84;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAA 64
           ||||||| |||||||||||||||||||||||||||||| |||||||||||||||||| |
Db     107 GTTGTGAAAATTCTTAATTCATGAGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGA 166

Qy      65 CTAAAGCAAGTGTTGGATTTAAAGCTGGTG 94
           |||||||||||||||||| |||||| | |
Db     167 CTAAAGCAAGTGTTGGATTCAAAGCTAGCG 196
```

```
RESULT 8
AAZ61365
ID   AAZ61365 standard; DNA; 195 BP.
XX
AC   AAZ61365;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Nucleotide sequence of chimeric promoter PrrnLrbcL+DBwt.
XX
KW   Chimeric promoter; plastid rRNA operon omega-type promoter; haemoglobin;
KW   Prrn promoter; rbcL; protein expression; vaccine; enzyme; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        7..89
FT                   /*tag=  a
FT                   /note= "Prrn plastid promoter"
FT   misc_signal     138..142
FT                   /*tag=  b
FT                   /note= "Shine-Dalgarno sequence"
XX
PN   WO200007431-A1.
XX
PD   17-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17806.
XX
PR   03-AUG-1998;   98US-0095163.
PR   03-AUG-1998;   98US-0095167.
PR   15-DEC-1998;   98US-0112257.
PR   29-APR-1999;   99US-0131611.
PR   11-JUN-1999;   99US-0138764.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
```

```
PI   Maliga P,  Kuroda H,  Khan MS;
XX
DR   WPI; 2000-205525/18.
XX
PT   New recombinant DNA constructs, for expressing high levels of
PT   heterologous protein in plastids of higher plants, includes promoter, a
PT   leader sequence and a downstream box element -
XX
PS   Claim 3; Fig 3B; 164pp; English.
XX
CC   The present sequence represents a chimeric plastid rRNA operon
CC   omega-type (Prrn) promoter with rbcL translation control sequences and
CC   a wild type downstream box. The chimeric promoter is used, as a 5'
CC   regulatory sequence, to produce recombinant DNA constructs for expressing
CC   heterologous proteins in the plastids of higher plants. The DNA
CC   constructs comprise a 5' regulatory region which includes a promoter
CC   element, a leader sequence and a downstream box element operably linked
CC   to a coding region of the heterologous protein. The chimeric regulatory
CC   region enhances translational efficiency of an mRNA molecule encoded by
CC   the DNA construct. The DNA constructs are used for producing transformed
CC   monocot and dicot plants having high levels of heterologous protein
CC   expression. They can be used to drive expression of proteins with
CC   agronomic, industrial or pharmaceutical importance, including production
CC   of vaccines, healthcare products like human haemoglobin, industrial or
CC   household enzymes. Plants which can be transformed with the constructs
CC   of the invention include maize, millet, sorghum, sugar cane, rice,
CC   wheat, barley, oat, rye or turf grass.
XX
SQ   Sequence 195 BP; 51 A; 33 C; 60 G; 51 T; 0 other;

  Query Match             19.0%;  Score 80;  DB 21;  Length 195;
  Best Local Similarity   94.3%;  Pred. No. 1.2e-14;
  Matches   83;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAA 64
           ||||||| |||||||||||||||||||||||||||||| |||||||||||||||||| |
Db     107 GTTGTGAAAATTCTTAATTCATGAGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGA 166

Qy      65 CTAAAGCAAGTGTTGGATTTAAAGCTGG 92
           |||||||||||||||||| |||||| |
Db     167 CTAAAGCAAGTGTTGGATTCAAAGCTAG 194


RESULT 9
AAZ61367
ID   AAZ61367 standard; DNA; 195 BP.
XX
AC   AAZ61367;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Nucleotide sequence of chimeric promoter PrrnLrbcL+DBm.
XX
KW   Chimeric promoter; plastid rRNA operon omega-type promoter; haemoglobin;
KW   Prrn promoter; rbcL; protein expression; vaccine; enzyme; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        7..89
FT                   /*tag=  a
FT                   /note= "Prrn plastid promoter"
FT   misc_signal     138..142
FT                   /*tag=  b
FT                   /note= "Shine-Dalgarno sequence"
FT   misc_RNA        153..175
FT                   /*tag=  c
XX
PN   WO200007431-A1.
XX
PD   17-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17806.
XX
PR   03-AUG-1998;   98US-0095163.
PR   03-AUG-1998;   98US-0095167.
PR   15-DEC-1998;   98US-0112257.
PR   29-APR-1999;   99US-0131611.
PR   11-JUN-1999;   99US-0138764.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Maliga P,  Kuroda H,  Khan MS;
XX
DR   WPI; 2000-205525/18.
XX
PT   New recombinant DNA constructs, for expressing high levels of
PT   heterologous protein in plastids of higher plants, includes promoter, a
PT   leader sequence and a downstream box element -
XX
PS   Claim 3; Fig 3B; 164pp; English.
XX
CC   The present sequence represents a chimeric plastid rRNA operon
CC   omega-type (Prrn) promoter with rbcL translation control sequences and
CC   a mutated downstream box. The chimeric promoter is used, as a 5'
CC   regulatory sequence, to produce recombinant DNA constructs for expressing
CC   heterologous proteins in the plastids of higher plants. The DNA
CC   constructs comprise a 5' regulatory region which includes a promoter
CC   element, a leader sequence and a downstream box element operably linked
CC   to a coding region of the heterologous protein. The chimeric regulatory
CC   region enhances translational efficiency of an mRNA molecule encoded by
CC   the DNA construct. The DNA constructs are used for producing transformed
CC   monocot and dicot plants having high levels of heterologous protein
CC   expression. They can be used to drive expression of proteins with
CC   agronomic, industrial or pharmaceutical importance, including production
CC   of vaccines, healthcare products like human haemoglobin, industrial or
CC   household enzymes. Plants which can be transformed with the constructs
CC   of the invention include maize, millet, sorghum, sugar cane, rice,
CC   wheat, barley, oat, rye or turf grass.
XX
SQ   Sequence 195 BP; 51 A; 34 C; 60 G; 47 T; 3 U; 0 other;

  Query Match             15.2%;  Score 64;  DB 21;  Length 195;
  Best Local Similarity   83.0%;  Pred. No. 1e-09;
  Matches   73;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAA 64
           ||||||| |||||||||||||||||||||||||||||| |||||   || || |||||||
Db     107 GTTGTGAAAATTCTTAATTCATGAGTTGTAGGGAGGGATTTATGAGUCCUCAGACAGAAA 166

Qy      65 CTAAAGCAAGTGTTGGATTTAAAGCTGG 92
           | |||||    || ||||| |||||| |
Db     167 CAAAAGCCUCAGTAGGATTCAAAGCTAG 194


RESULT 10
AAZ61366
ID   AAZ61366 standard; DNA; 159 BP.
XX
AC   AAZ61366;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Nucleotide sequence of chimeric promoter PrrnLrbcL-DB.
XX
KW   Chimeric promoter; plastid rRNA operon omega-type promoter; haemoglobin;
KW   Prrn promoter; rbcL; protein expression; vaccine; enzyme; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        7..89
FT                   /*tag=  a
FT                   /note= "Prrn plastid promoter"
FT   misc_signal     138..142
FT                   /*tag=  b
```

```
FT                   /note= "Shine-Dalgarno sequence"
XX
PN   WO200007431-A1.
XX
PD   17-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17806.
XX
PR   03-AUG-1998;   98US-0095163.
PR   03-AUG-1998;   98US-0095167.
PR   15-DEC-1998;   98US-0112257.
PR   29-APR-1999;   99US-0131611.
PR   11-JUN-1999;   99US-0138764.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Maliga P,  Kuroda H,  Khan MS;
XX
DR   WPI; 2000-205525/18.
XX
PT   New recombinant DNA constructs, for expressing high levels of
PT   heterologous protein in plastids of higher plants, includes promoter, a
PT   leader sequence and a downstream box element -
XX
PS   Claim 3; Fig 3B; 164pp; English.
XX
CC   The present sequence represents a chimeric plastid rRNA operon
CC   omega-type (Prrn) promoter with rbcL translation control sequences,
CC   without a downstream box. The chimeric promoter is used, as a 5'
CC   regulatory sequence, to produce recombinant DNA constructs for expressing
CC   heterologous proteins in the plastids of higher plants. The DNA
CC   constructs comprise a 5' regulatory region which includes a promoter
CC   element, a leader sequence and a downstream box element operably linked
CC   to a coding region of the heterologous protein. The chimeric regulatory
CC   region enhances translational efficiency of an mRNA molecule encoded by
CC   the DNA construct. The DNA constructs are used for producing transformed
CC   monocot and dicot plants having high levels of heterologous protein
CC   expression. They can be used to drive expression of proteins with
CC   agronomic, industrial or pharmaceutical importance, including production
CC   of vaccines, healthcare products like human haemoglobin, industrial or
CC   household enzymes. Plants which can be transformed with the constructs
CC   of the invention include maize, millet, sorghum, sugar cane, rice,
CC   wheat, barley, oat, rye or turf grass.
XX
SQ   Sequence 159 BP; 35 A; 26 C; 53 G; 45 T; 0 other;

  Query Match             10.5%;  Score 44.2;  DB 21;  Length 159;
  Best Local Similarity   93.9%;  Pred. No. 0.0012;
  Matches   46;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACC 53
           ||||||| |||||||||||||||||||||||||||||| |||||||| |
Db     107 GTTGTGAAAATTCTTAATTCATGAGTTGTAGGGAGGGATTTATGTCAGC 155


RESULT 11
AAX21432
ID   AAX21432 standard; DNA; 1416 BP.
XX
AC   AAX21432;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Regulatory region Prrn(L)/rbcL(S)/kan/TpsbA(L).
XX
KW   Construct; marker; antibiotic resistance; regulatory sequence; promoter;
KW   stabilizing sequence; plastid; plant; ss.
XX
OS   Synthetic.
XX
PN   US5877402-A.
XX
PD   02-MAR-1999.
XX
PF   31-JAN-1994;   94US-0189256.
XX
PR   31-JAN-1994;   94US-0189256.
PR   01-MAY-1990;   90US-0518763.
PR   25-AUG-1993;   93US-0111398.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Allison LA,  Carrer H,  Kanevski I,  Maliga P,  Maliga ZS;
PI   Staub JM,  Zoubenko O;
XX
DR   WPI; 1999-189744/16.
XX
PT   New DNA construct containing an antibiotic resistance marker -
PT   useful for stably transforming the plastids of multicellular plants
XX
PS   Disclosure; Fig 28A; 101pp; English.
XX
CC   The invention relates to a DNA construct, containing a non-lethal stable
CC   marker (e.g antibiotic resistance) under the control of a 5' regulatory
CC   sequence and a 3' stabilizing sequence, for stably transforming the
CC   plastids of multicellular plants and allowing expression of heterologous
CC   proteins especially non-native plastid or plant proteins. This sequence
CC   corresponds to the regulatory region Prrn(L)/rbcL(S)/kan/TpsbA(L) from
CC   the plasmid pTNH7.
XX
SQ   Sequence 1416 BP; 312 A; 322 C; 379 G; 403 T; 0 other;

  Query Match             9.8%;  Score 41;  DB 20;  Length 1416;
  Best Local Similarity   72.6%;  Pred. No. 0.028;
  Matches   53;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAA 87
           ||||||||||||||| |||||||||||||||||||    |  | |   | ||||||  |
Db     139 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGATTGAACAAGATGGATTGCAC 198

Qy      88 GCTGGTGTTAAGG 100
           || |||  |  ||
Db     199 GCAGGTTCTCCGG 211


RESULT 12
AAT85196
ID   AAT85196 standard; DNA; 1417 BP.
XX
AC   AAT85196;
XX
DT   10-MAR-1998  (first entry)
XX
DE   Plastid targeting region of plasmid pGS85A.
XX
KW   Plastid; transformation; transplastomic plant; transgenic plant;
KW   Brassica; cruciferous plant; vector; plasmid pGS85A;
KW   kanamycin resistance; neomycin phosphotransferase; neo gene;
KW   selectable marker; ss.
XX
OS   Chimeric - Arabidopsis thaliana.
OS   Chimeric - Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             158..970
FT                   /*tag=  a
FT                   /note= "kanamycin resistance gene"
XX
PN   WO9732977-A1.
XX
PD   12-SEP-1997.
XX
PF   06-MAR-1997;   97WO-US03444.
XX
PR   06-MAR-1996;   96US-0012916.
XX
```

```
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Maliga P,  Reddy SV,  Sikdar SR;
XX
DR   WPI; 1997-470543/43.
XX
PT   Production of transplastomic plants - by transfection with plastid
PT   targetted DNA, specifically for Cruciferous plants
XX
PS   Disclosure; Fig 7; 45pp; English.
XX
CC   This sequence comprises the plastid targeting region of plasmid
CC   pGS85A.  This plasmid carries a neomycin phosphotransferase (neo)
CC   gene that confers kanamycin resistance gene, flanked on both sides
CC   by Arabidopsis plastid targeting sequences.  It was produced by
CC   ligating the neo gene into the unique HincII site of plasmid pGS7
CC   (see AAT85194).  A novel method of producing transplastomic plants
CC   involves: delivering transforming DNA, preferably cloned in pGS31A,
CC   pGS85A or pGS7, to a plastid genome of plant cells in culture,
CC   selecting cells with transformed plastids and regenerating these to
CC   plants.  Transforming DNA comprises (i) (i) several targetting
CC   sequences (i.e. plastid DNA sequences from the genome to be
CC   transformed) to allow homologous recombination, (ii) 5' and 3'
CC   regulatory regions from plastid DNA linked to a selectable marker
CC   gene, (iii) regulatory regions controlling a foreign gene of
CC   interest, and (iv) at a cloning site for insertion of the foreign
CC   gene adjacent to the marker gene.  The method is used to transform
CC   cruciferous plants, e.g. Brassica species.  The combination of
CC   homologous recombination, selectable marker and use of plastid
CC   regulatory sequences improves generation of stably transformed
CC   plants.  Expression of foreign genes in plastids, rather than in
CC   the nucleus, avoids risk of transmitting the foreign gene in the
CC   pollen, provides high levels of protein expression, permits
CC   simultaneous incorporation of several genes and avoids the
CC   positional effects and gene silencing associated with nuclear
CC   transformation.
XX
SQ   Sequence 1417 BP; 312 A; 323 C; 379 G; 403 T; 0 other;

  Query Match             9.8%;  Score 41;  DB 18;  Length 1417;
  Best Local Similarity   72.6%;  Pred. No. 0.028;
  Matches   53;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAA 87
           |||||||||||||||| ||||||||||||||||||    | | |   | ||||||  |
Db     140 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGATTGAACAAGATGGATTGCAC 199

Qy      88 GCTGGTGTTAAGG 100
           || |||  |  ||
Db     200 GCAGGTTCTCCGG 212


RESULT 13
AAX21420
ID   AAX21420 standard; DNA; 278 BP.
XX
AC   AAX21420;
XX
DT   21-MAY-1999  (first entry)
XX
DE   5' and 3' regulatory region for plasmid PrbcL.
XX
KW   Construct; marker; antibiotic resistance; regulatory sequence; promoter;
KW   stabilizing sequence; plastid; plant; ss.
XX
OS   Synthetic.
XX
PN   US5877402-A.
XX
PD   02-MAR-1999.
XX
PF   31-JAN-1994;   94US-0189256.
XX
PR   31-JAN-1994;   94US-0189256.
PR   01-MAY-1990;   90US-0518763.
PR   25-AUG-1993;   93US-0111398.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Allison LA,  Carrer H,  Kanevski I,  Maliga P,  Maliga ZS;
PI   Staub JM,  Zoubenko O;
XX
DR   WPI; 1999-189744/16.
XX
PT   New DNA construct containing an antibiotic resistance marker -
PT   useful for stably transforming the plastids of multicellular plants
XX
PS   Disclosure; Fig 23C; 101pp; English.
XX
CC   The invention relates to a DNA construct, containing a non-lethal stable
CC   marker (e.g antibiotic resistance) under the control of a 5' regulatory
CC   sequence and a 3' stabilizing sequence, for stably transforming the
CC   plastids of multicellular plants and allowing expression of heterologous
CC   proteins especially non-native plastid or plant proteins. This sequence
CC   corresponds to the regulatory region for the plasmid PrbcL.
XX
SQ   Sequence 278 BP; 102 A; 30 C; 56 G; 90 T; 0 other;

  Query Match             9.3%;  Score 39.2;  DB 20;  Length 278;
  Best Local Similarity   93.2%;  Pred. No. 0.05;
  Matches   41;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATG 48
           |||||||||||||||||||||||||||||||||||||||||   |||
Db     234 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGATCCATG 277


RESULT 14
AAX13041
ID   AAX13041 standard; DNA; 6433 BP.
XX
AC   AAX13041;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:104.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 672-676; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
```

```
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 6433 BP; 1816 A; 1501 C; 1085 G; 2022 T; 9 other;

  Query Match             9.0%;  Score 38;  DB 20;  Length 6433;
  Best Local Similarity   46.0%;  Pred. No. 0.43;
  Matches  128;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy       3 GCGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGA 62
           ||| | | | ||||| |      | | |||| |      | ||    || | |   ||
Db    3674 GCGCTTTAAAAATTCATCCAAGGTCATTTGTTGACGCCAATTTGAAGCAACCCGCGCATG 3733

Qy      63 AACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACAC 122
             ||  | |||         ||  | | |||||  | |||||   |||   ||   |  ||
Db    3734 CTCTTTAACAATGACACCTTTACAGGTTGGTGCAATGGATTCATAATCATCTCGGTTGAC 3793

Qy     123 CCCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCT 182
           |||| | | |  ||| || |     |  |   |       |  | | |||  | |  |
Db    3794 CCCGTAATTCCCAACTAATGGATACGTAAAAGTAATCATTTGTCCATTAAAACTTTGGTC 3853

Qy     183 CGGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATG 242
            |  || |  ||   || | |      ||||||       | ||||| ||  |||||
Db    3854 AGTAATTGCTTCTTGGTAACCTGTCATGCCAGTAGTAAAAACAACTTCTCCTACTACATT 3913

Qy     243 GACAACTGTTTGGACTGATGGACTTACCAGTCTTGATC 280
           | |  |||     |  | |   | | | | |  || ||
Db    3914 GCCTTCTGCACCAAACGCTTTTCCTTCAAATACTGTTC 3951


RESULT 15
AAH41225/c
ID   AAH41225 standard; DNA; 349980 BP.
XX
AC   AAH41225;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Pyrococcus abyssi genomic fragment #4.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
OS   Pyrococcus abyssi.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..49980
FT                   /*tag=  a
FT                   /note= "This sequence overlaps with the 3' end of
FT                   AAH41224"
FT   misc_feature    300001..349980
FT                   /*tag=  b
FT                   /note= "This sequence overlaps with the 5'end of
FT                   AAH41226"
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 1; Page 429-524; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC   hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC   vents. The present sequence is a fragment of the genomic sequence of P.
CC   abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41224
CC   and the 3' end of this sequence overlaps with the 5' end of AAH41226. The
CC   proteins of the present invention have various potential industrial uses,
CC   since the proteins are stable at very high temperatures, some up to 110
CC   degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence 349980 BP; 99421 A; 76616 C; 77444 G; 96499 T; 0 other;

  Query Match             8.7%;  Score 36.6;  DB 22;  Length 349980;
  Best Local Similarity   60.6%;  Pred. No. 6.1;
  Matches   60;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     321 GGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCTATTTGAAGAGGGTTC 380
           | ||   ||    |||||     |  ||||||| ||  ||  ||||||||||||||||
Db 268563 GCACGGCGAGGGTTATATAGCAAAGATAGCTTACCCCCTAACCCTATTTGAAGAGGGAAG 268504

Qy     381 TTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATT 419
           |    | |  |||||| |   || |  || ||||||||
Db 268503 CTTAGTTCAGCTGTTTAGTGCTATAGCTGGGAACGTATT 268465


Search completed: April 16, 2003, 02:32:58
Job time : 4060 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 11:36:59 ; Search time 91 Seconds
(without alignments)
1415.431 Million cell updates/sec

Title: US-09-696-664A-3
Perfect score: 420
Sequence: 1 gcgcgttgtgagaattctta.....tcattgtgggtaacgtattt 420

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

[redacted]ched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
[redacted]	1	41	9.8	1416	2	US-08-189-256A-27	Sequence 27, Appl
	2	41	9.8	1416	4	US-09-193-853-27	Sequence 27, Appl
	3	41	9.8	1417	4	US-09-142-114B-7	Sequence 7, Appli
	4	39.2	9.3	278	2	US-08-189-256A-15	Sequence 15, Appl
	5	39.2	9.3	278	4	US-09-193-853-15	Sequence 15, Appl
	6	36.2	8.6	191	2	US-08-189-256A-5	Sequence 5, Appli
	7	36.2	8.6	191	4	US-09-193-853-5	Sequence 5, Appli
	8	34.4	8.2	165	2	US-08-189-256A-4	Sequence 4, Appli
	9	34.4	8.2	165	4	US-09-193-853-4	Sequence 4, Appli
	10	34.2	8.1	3381	4	US-09-336-447A-6	Sequence 6, Appli
c	11	32.8	7.8	52	2	US-08-189-256A-35	Sequence 35, Appl
c	12	32.8	7.8	52	4	US-09-193-853-35	Sequence 35, Appl
	13	31.8	7.6	168	2	US-08-189-256A-2	Sequence 2, Appli
	14	31.8	7.6	168	4	US-09-193-853-2	Sequence 2, Appli
	15	31.8	7.6	3995	4	US-09-428-711A-13	Sequence 13, Appl
	16	31.6	7.5	1704	1	US-08-460-934-5	Sequence 5, Appli
	17	31.6	7.5	1704	2	US-08-782-118-5	Sequence 5, Appli
	18	31.6	7.5	4429	1	US-08-308-872B-3	Sequence 3, Appli
	19	31.6	7.5	4435	1	US-08-308-872B-5	Sequence 5, Appli
c	20	31.6	7.5	8075	3	US-08-374-077C-1	Sequence 1, Appli
c	21	31.6	7.5	8075	4	US-08-895-590-1	Sequence 1, Appli
c	22	31.6	7.5	8075	4	US-09-539-879A-1	Sequence 1, Appli
	23	31.2	7.4	329	4	US-08-991-789A-251	Sequence 251, App
	24	31.2	7.4	329	4	US-09-062-451-251	Sequence 251, App
	25	30	7.1	8439	4	US-09-221-017B-473	Sequence 473, App
	26	29.8	7.1	672	4	US-09-080-983-12	Sequence 12, Appl
c	27	29.8	7.1	7176	4	US-09-221-017B-626	Sequence 626, App
	28	29.8	7.1	7812	4	US-09-368-590-1	Sequence 1, Appli
	29	29.8	7.1	15500	4	US-09-080-983-1	Sequence 1, Appli
c	30	29.8	7.1	38844	4	US-09-734-675-3	Sequence 3, Appli
	31	29.6	7.0	414	4	US-09-042-353-353	Sequence 353, App
	32	29.6	7.0	414	4	US-08-758-417A-201	Sequence 201, App
c	33	29.4	7.0	379	1	US-08-145-617-5	Sequence 5, Appli
c	34	29.4	7.0	2190	4	US-09-625-188-19	Sequence 19, Appl
	35	29.2	7.0	5398	3	US-09-356-952-11	Sequence 11, Appl
	36	29	6.9	1727	4	US-09-071-035-295	Sequence 295, App
	37	29	6.9	1839	4	US-09-071-035-293	Sequence 293, App
c	38	28.8	6.9	499	4	US-09-004-838-9	Sequence 9, Appli
c	39	28.8	6.9	1110	1	US-08-257-341-4	Sequence 4, Appli
	40	28.6	6.8	701	3	US-08-331-625A-55	Sequence 55, Appl
	41	28.6	6.8	701	4	US-09-494-151-55	Sequence 55, Appl
	42	28.6	6.8	777	4	US-08-809-326A-7	Sequence 7, Appli
	43	28.6	6.8	813	4	US-08-809-326A-4	Sequence 4, Appli
	44	28.6	6.8	1048	4	US-08-809-326A-9	Sequence 9, Appli
	45	28.6	6.8	1284	3	US-08-331-625A-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-08-189-256A-27
; Sequence 27, Application US/08189256A
; Patent No. 5877402
; GENERAL INFORMATION:
; APPLICANT: Maliga, Pal
; APPLICANT: Svab, Zora
; APPLICANT: Staub, Jeffrey
; APPLICANT: Zoubenko, Oleg V.
; APPLICANT: Allison, Lori A.
; APPLICANT: Carrer, Helaine
; APPLICANT: Kanevski, Ivan
; TITLE OF INVENTION: DNA Constructs and Methods for Stably
; TITLE OF INVENTION: Transforming Plastids of Multicellular Plants and
; TITLE OF INVENTION: Expressing Recombinant Proteins Therein
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/189,256A
; FILING DATE: 31-JAN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/111,398
; FILING DATE: 25-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US o7/518,763
; FILING DATE: 01-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-189-256A-27

  Query Match              9.8%;  Score 41;  DB 2;  Length 1416;
  Best Local Similarity   72.6%;  Pred. No. 0.0013;
  Matches   53;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAA 87
           ||||||||||||||| |||||||||||||||||||    |  | |   | ||||||  |
Db     139 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGATTGAACAAGATGGATTGCAC 198

Qy      88 GCTGGTGTTAAGG 100
           || |||  |  ||
Db     199 GCAGGTTCTCCGG 211


RESULT 2
US-09-193-853-27
; Sequence 27, Application US/09193853
; Patent No. 6388168
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/193,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/189,256
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1416 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-193-853-27

  Query Match              9.8%;  Score 41;  DB 4;  Length 1416;
  Best Local Similarity   72.6%;  Pred. No. 0.0013;
  Matches   53;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAA 87
           ||||||||||||||| |||||||||||||||||||    |  | |   | ||||||  |
Db     139 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGATTGAACAAGATGGATTGCAC 198

Qy      88 GCTGGTGTTAAGG 100
           || |||  |  ||
Db     199 GCAGGTTCTCCGG 211


RESULT 3
US-09-142-114B-7
; Sequence 7, Application US/09142114B
; Patent No. 6376744
; GENERAL INFORMATION:
;  APPLICANT:  Rutgers University
;  APPLICANT:  Maliga, Pal
;  APPLICANT:  Sikdar, Samir R.
;  APPLICANT:  Reddy, Siva Vanga
;  TITLE OF INVENTION:  Plastid Transformation in Arabidopsis
;  TITLE OF INVENTION:  Thaliana
;  FILE REFERENCE:  09/142,114
;  CURRENT APPLICATION NUMBER: US/09/142,114B
;  CURRENT FILING DATE:  1999-02-05
;  PRIOR APPLICATION NUMBER:  PCT/US97/03444
;  PRIOR FILING DATE:  1997-03-06
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1417
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Sequence source:/note="synthetic construct"
; Patent No. 6376744
US-09-142-114B-7

  Query Match              9.8%;  Score 41;  DB 4;  Length 1417;
  Best Local Similarity   72.6%;  Pred. No. 0.0013;
  Matches   53;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAA 87
           ||||||||||||||| |||||||||||||||||||    |  | |   | ||||||  |
Db     140 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGATTGAACAAGATGGATTGCAC 199

Qy      88 GCTGGTGTTAAGG 100
           || |||  |  ||
Db     200 GCAGGTTCTCCGG 212


RESULT 4
US-08-189-256A-15
; Sequence 15, Application US/08189256A
; Patent No. 5877402
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
```

```
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/189,256A
;      FILING DATE:  31-JAN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/111,398
;      FILING DATE:  25-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  278 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-189-256A-15

  Query Match             9.3%;  Score 39.2;  DB 2;  Length 278;
  Best Local Similarity   93.2%;  Pred. No. 0.0024;
  Matches   41;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATG 48
           |||||||||||||||||||||||||||||||||||||   |||
Db     234 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGATCCATG 277


RESULT 5
US-09-193-853-15
; Sequence 15, Application US/09193853
; Patent No. 6388168
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/193,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/189,256
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  278 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-193-853-15

  Query Match             9.3%;  Score 39.2;  DB 4;  Length 278;
  Best Local Similarity   93.2%;  Pred. No. 0.0024;
  Matches   41;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATG 48
           |||||||||||||||||||||||||||||||||||||   |||
Db     234 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGATCCATG 277


RESULT 6
US-08-189-256A-5
; Sequence 5, Application US/08189256A
; Patent No. 5877402
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/189,256A
;      FILING DATE:  31-JAN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/111,398
;      FILING DATE:  25-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  191 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-189-256A-5

  Query Match             8.6%;  Score 36.2;  DB 2;  Length 191;
  Best Local Similarity   83.7%;  Pred. No. 0.019;
  Matches  41;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTG 76
           ||||||||||||||| |||||||||||||||||||     |||| |||
Db     143 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGGAAGCGGGTG 191


RESULT 7
US-09-193-853-5
; Sequence 5, Application US/09193853
; Patent No. 6388168
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/193,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/189,256
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  191 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-193-853-5

  Query Match             8.6%;  Score 36.2;  DB 4;  Length 191;
  Best Local Similarity   83.7%;  Pred. No. 0.019;
  Matches  41;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTG 76
           ||||||||||||||| |||||||||||||||||||     |||| |||
Db     143 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGGAAGCGGGTG 191


RESULT 8
US-08-189-256A-4
; Sequence 4, Application US/08189256A
; Patent No. 5877402
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/189,256A
;      FILING DATE:  31-JAN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/111,398
;      FILING DATE:  25-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  165 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-189-256A-4

  Query Match             8.2%;  Score 34.4;  DB 2;  Length 165;
```

```
  Best Local Similarity   86.4%;  Pred. No. 0.067;
  Matches   38;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGC 71
           ||||||||||||||| ||||||||||||||||||     ||||
Db     118 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGGAAGC 161


RESULT 9
US-09-193-853-4
; Sequence 4, Application US/09193853
; Patent No. 6388168
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/193,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/189,256
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  165 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-193-853-4

  Query Match              8.2%;  Score 34.4;  DB 4;  Length 165;
  Best Local Similarity   86.4%;  Pred. No. 0.067;
  Matches   38;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGC 71
           ||||||||||||||| ||||||||||||||||||     ||||
Db     118 AGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGGAAGC 161


RESULT 10
US-09-336-447A-6
; Sequence 6, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
;  APPLICANT: HANSEN, ERIC J.
;  APPLICANT:  AEBI, CHRISTOPH
;  APPLICANT:  COPE, LESLIE D.
;  APPLICANT:  MACIVER, ISOBEL
;  APPLICANT:  FISKE, MICHAEL J.
;  APPLICANT:  FREDENBURG, ROSS A.
;  TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
;  FILE REFERENCE: AMCY:024
;  CURRENT APPLICATION NUMBER: US/09/336,447A
;  CURRENT FILING DATE:  1999-06-21
;  NUMBER OF SEQ ID NOS: 98
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 3381
;   TYPE: DNA
;   ORGANISM: Moraxella catarrhalis
US-09-336-447A-6

  Query Match              8.1%;  Score 34.2;  DB 4;  Length 3381;
  Best Local Similarity   49.7%;  Pred. No. 0.32;
  Matches   87;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy      61 GAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTAC 120
           | |  |||    |||| ||  |  | || || |||||| |  |   |  |   |  |
Db    1272 GCAGGTAATATCAGTGCTGATTCAACAGATGCTGTTAATGGCTCACAGCTATATGCTTTG 1331

Qy     121 ACCCCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAG 180
           || | |    | || || | |    ||| | |||   |||| ||     |||||   |
Db    1332 GCCACAGCTGTCAAAGCCGATGCCGATGAAAACTTTAAAGCACTCACCAAAACTCAAAAT 1391

Qy     181 CTCGGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTG 235
               | ||  ||  |  |||||| ||  ||| ||  ||  ||  |  ||||
Db    1392 ACTTTGATTGAGCAAGGTGAAGCACAAGACGCATTAATCGCTCAAAATCAAACTG 1446


RESULT 11
US-08-189-256A-35/c
; Sequence 35, Application US/08189256A
; Patent No. 5877402
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/189,256A
;      FILING DATE:  31-JAN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/111,398
```

```
;      FILING DATE:  25-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  52 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-189-256A-35

  Query Match             7.8%;  Score 32.8;  DB 2;  Length 52;
  Best Local Similarity   76.9%;  Pred. No. 0.13;
  Matches   40;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      29 GTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGG 80
           ||||||||||||| |||||||||||||||||||    |  | |   | |||
Db      52 GTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGATTGAACAAGATGG 1


RESULT 12
US-09-193-853-35/c
; Sequence 35, Application US/09193853
; Patent No. 6388168
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/193,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/189,256
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  52 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-193-853-35

  Query Match             7.8%;  Score 32.8;  DB 4;  Length 52;
  Best Local Similarity   76.9%;  Pred. No. 0.13;
  Matches   40;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      29 GTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGG 80
           ||||||||||||| |||||||||||||||||||    |  | |   | |||
Db      52 GTTGTAGGGAGGGATTTATGTCACCACAAACAGAGGGGATTGAACAAGATGG 1


RESULT 13
US-08-189-256A-2
; Sequence 2, Application US/08189256A
; Patent No. 5877402
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/189,256A
;      FILING DATE:  31-JAN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/111,398
;      FILING DATE:  25-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  168 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
US-08-189-256A-2

  Query Match             7.6%;  Score 31.8;  DB 2;  Length 168;
  Best Local Similarity   94.3%;  Pred. No. 0.47;
  Matches   33;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGA 62
           ||||||||||||||| |||| |||||||||||||||
Db     118 AGTTGTAGGGAGGGATTTATCTCACCACAAACAGA 152


RESULT 14
US-09-193-853-2
; Sequence 2, Application US/09193853
; Patent No. 6388168
;  GENERAL INFORMATION:
;    APPLICANT:  Maliga, Pal
;    APPLICANT:  Svab, Zora
;    APPLICANT:  Staub, Jeffrey
;    APPLICANT:  Zoubenko, Oleg V.
;    APPLICANT:  Allison, Lori A.
;    APPLICANT:  Carrer, Helaine
;    APPLICANT:  Kanevski, Ivan
;    TITLE OF INVENTION:  DNA Constructs and Methods for Stably
;    TITLE OF INVENTION:  Transforming Plastids of Multicellular Plants and
;    TITLE OF INVENTION:  Expressing Recombinant Proteins Therein
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street  Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/193,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/189,256
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US o7/518,763
;      FILING DATE:  01-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  168 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-193-853-2

  Query Match             7.6%;  Score 31.8;  DB 4;  Length 168;
  Best Local Similarity   94.3%;  Pred. No. 0.47;
  Matches   33;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      28 AGTTGTAGGGAGGGACTTATGTCACCACAAACAGA 62
           ||||||||||||||| |||| |||||||||||||||
Db     118 AGTTGTAGGGAGGGATTTATCTCACCACAAACAGA 152


RESULT 15
US-09-428-711A-13
; Sequence 13, Application US/09428711A
; Patent No. 6358720
; GENERAL INFORMATION:
;  APPLICANT:  Muramatsu, Masaaki
;  APPLICANT:  Shirasawa, Takuji
;  APPLICANT:  Tokumitsu, Hiroshi
;  APPLICANT:  No. 6358720uchi, Teruhisa
;  TITLE OF INVENTION:  SERINE/THREONINE PROTEIN KINASE
;  FILE REFERENCE:  06501-045001
;  CURRENT APPLICATION NUMBER: US/09/428,711A
;  CURRENT FILING DATE:  1999-10-28
;  PRIOR APPLICATION NUMBER:  PCT/JP98/01246
;  PRIOR FILING DATE:  1998-03-23
;  PRIOR APPLICATION NUMBER:  JP 9/124798
;  PRIOR FILING DATE:  1997-04-28
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 3995
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (388)...(3540)
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(3995)
;   OTHER INFORMATION: n = A,T,C or G
US-09-428-711A-13

  Query Match             7.6%;  Score 31.8;  DB 4;  Length 3995;
  Best Local Similarity   51.8%;  Pred. No. 2;
  Matches   72;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     119 ACACCCCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTC 178
           ||| ||   || |  || | | ||| | || |  ||  || |  ||  ||| ||||  |
Db    2537 ACAGCCTACAGGAGAAACCTATGGAGATTGCTCCCTCTGCTGGCTTTGGAGGGACTCTGC 2596

Qy     179 AGCTCGGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTA 238
           | |  || | ||    || ||  ||  |     ||| | ||| || || | ||   ||
Db    2597 ATCCAGGAGCTCGTGGTGGAGGGGCCAGCAGCCCAGCACCTGTGGTATTTACTGTAGGCT 2656

Qy     239 CATGGACAACTGTTTGGAC 257
           |     | | || |   ||
Db    2657 CCCCACCCAGTGGTGCCAC 2675


Search completed: April 15, 2003, 18:27:33
Job time : 97 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 10:08:49 ; Search time 85 Seconds
(without alignments)
4334.241 Million cell updates/sec

Title: US-09-696-664A-3
Perfect score: 420
Sequence: 1 gcgcgttgtgagaattctta.....tcattgtgggtaacgtattt 420

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	336.6	80.1	2004	10	US-09-887-576-160	Sequence 160, App
	2	250.8	59.7	406	10	US-09-878-574-3601	Sequence 3601, Ap
	3	38	9.0	6433	10	US-09-070-927A-104	Sequence 104, App
c	4	34.8	8.3	22008	9	US-10-091-504-2110	Sequence 2110, Ap
c	5	34.8	8.3	22008	10	US-09-764-869-2110	Sequence 2110, Ap
	6	34.4	8.2	42000	9	US-10-081-563-25	Sequence 25, Appl
	7	34.4	8.2	465237	10	US-09-933-267A-1	Sequence 1, Appli
	8	34.2	8.1	3381	9	US-09-952-267-6	Sequence 6, Appli
c	9	32.8	7.8	495	9	US-10-092-154-1228	Sequence 1228, Ap
c	10	32.8	7.8	495	9	US-10-092-154-1229	Sequence 1229, Ap
c	11	32.8	7.8	495	10	US-09-764-847-1228	Sequence 1228, Ap
c	12	32.8	7.8	495	10	US-09-764-847-1229	Sequence 1229, Ap
	13	32.6	7.8	2373	9	US-09-738-626-1030	Sequence 1030, Ap
	14	31.8	7.6	2406	9	US-09-938-842A-2369	Sequence 2369, Ap
	15	31.2	7.4	329	9	US-09-924-400-251	Sequence 251, App
	16	31.2	7.4	329	10	US-09-810-936-251	Sequence 251, App
	17	31.2	7.4	329	10	US-09-429-755-251	Sequence 251, App
c	18	31	7.4	1284	9	US-09-938-842A-2349	Sequence 2349, Ap
	19	30.8	7.3	1530	9	US-09-738-626-2259	Sequence 2259, Ap
	20	30.6	7.3	370	10	US-09-960-352-6403	Sequence 6403, Ap
c	21	30.4	7.2	468	10	US-09-880-107-3655	Sequence 3655, Ap
	22	30.4	7.2	498	10	US-09-880-107-140	Sequence 140, App
	23	30.4	7.2	3115	9	US-09-981-353-85	Sequence 85, Appl
c	24	30.4	7.2	3115	9	US-09-981-353-85	Sequence 85, Appl
c	25	30	7.1	411	9	US-10-040-739-599	Sequence 599, App
c	26	30	7.1	415	9	US-09-796-692-4492	Sequence 4492, Ap
c	27	30	7.1	649	9	US-10-184-644-384	Sequence 384, App
c	28	30	7.1	14800	10	US-09-954-456-1601	Sequence 1601, Ap
c	29	29.8	7.1	1472	7	US-08-781-986A-555	Sequence 555, App
c	30	29.8	7.1	9062	7	US-08-781-986A-85	Sequence 85, Appl
	31	29.8	7.1	21222	10	US-09-734-676-3	Sequence 3, Appli
c	32	29.8	7.1	38844	12	US-10-060-333-3	Sequence 3, Appli
c	33	29.6	7.0	293	10	US-09-864-761-18923	Sequence 18923, A
c	34	29.6	7.0	459	10	US-09-864-761-2182	Sequence 2182, Ap
c	35	29.6	7.0	170834	10	US-09-835-232-7	Sequence 7, Appli
c	36	29.4	7.0	8002	10	US-09-935-943-10	Sequence 10, Appl
	37	29.4	7.0	1503841	9	US-09-946-807-1	Sequence 1, Appli
	38	29.4	7.0	1503841	10	US-09-795-668-1	Sequence 1, Appli
	39	29.4	7.0	1503841	10	US-09-795-686-1	Sequence 1, Appli
c	40	29.2	7.0	316	10	US-09-864-761-17473	Sequence 17473, A
c	41	29.2	7.0	436	10	US-09-864-761-686	Sequence 686, App
	42	29.2	7.0	4770	10	US-09-801-368-63	Sequence 63, Appl
	43	29	6.9	7657	10	US-09-070-927A-33	Sequence 33, Appl
	44	28.8	6.9	254	9	US-09-796-692-4835	Sequence 4835, Ap
	45	28.8	6.9	254	9	US-09-796-692-6076	Sequence 6076, Ap

ALIGNMENTS

RESULT 1
US-09-887-576-160/c
; Sequence 160, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 160
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-160

Query Match 80.1%; Score 336.6; DB 10; Length 2004;
Best Local Similarity 88.2%; Pred. No. 4.3e-97;
Matches 366; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy       6 TTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAAC 65
           |||  |||||||| ||||||||| ||||||||||||||||||||||||||||||||| ||
Db    1239 TTGCAAGAATTCTAAATTCATGACTTGTAGGGAGGGACTTATGTCACCACAAACAGAGAC 1180

Qy      66 TAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCC 125
           |||||||||||||| || |||||||||||||| || |||||||||||||||| || ||
Db    1179 TAAAGCAAGTGTTGGGTTCAAAGCTGGTGTTAAAGAGTATAAATTGACTTACTATACTCC 1120
```

```
Qy    126 GGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGG 185
Db   1119 TGAATATGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACCTGG 1060

Qy    186 GGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGAC 245
Db   1059 AGTTCCACCTGAAGAAGCAGGGGCTGCGGTAGCTGCTGAATCTTCTACTGGTACATGGAC 1000

Qy    246 AACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACAT 305
Db    999 AACTGTGTGGACCGATGGGCTTACCAGCCTTGATCGTTACAAAGGACGATGCTACCACAT 940

Qy    306 CGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCT 365
Db    939 CGAGCCCGTTCCAGGAGAAGAAACTCAATTTATTGCGTATGTAGCTTATCCCTTAGACCT 880

Qy    366 ATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
Db    879 TTTTGAAGAAGGTTCGGTTACTAACATGTTTACCTCGATTGTGGGTAATGTATTT 825
```

RESULT 2
[redacted]09-878-574-3601
; Sequence 3601, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3601
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(406)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-008-Q1-B1-C9
US-09-878-574-3601

Query Match 59.7%; Score 250.8; DB 10; Length 406;
Best Local Similarity 90.5%; Pred. No. 4.3e-70;
Matches 267; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy      5 GTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAA 64
Db    111 GTTTCGAGAATTCTTAATTCATGAGTTGTAGGGAGGGATTTATGTCACCACAAACAGAGA 170

Qy     65 CTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCC 124
Db    171 CTAAAGCAAGTGTTGGGTTCAAAGCTGGTGTTAAAGATTATAAATTGACTTATTATACTC 230

Qy    125 CGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCG 184
Db    231 CTGACTATGAAACCAAAGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACCAG 290

Qy    185 GGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGA 244
Db    291 GAGTTCCGCCTGAAGAAGCAGGTGCCGCGGTAGCCGCCGAATCTTCTACTGGTACATGGA 350

Qy    245 CAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTA 299
Db    351 CAACTGTGTGGACCGATGGGCTTACCAGTCNTGATCGTTACAAAGGGCGATGCTA 405
```

RESULT 3
US-09-070-927A-104
; Sequence 104, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6433 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-09-070-927A-104

Query Match 9.0%; Score 38; DB 10; Length 6433;
Best Local Similarity 46.0%; Pred. No. 0.18;
Matches 128; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
Qy      3 GCGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGA 62
Db   3674 GCGCTTTAAAAATTCATCCAAGGTCATTTGTTGACGCCAATTTGAAGCAACCCGCGCATG 3733

Qy     63 AACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACAC 122
Db   3734 CTCTTTAACAATGACACCTTTACAGGTTGGTGCAATGGATTCATAATCATCTCGGTTGAC 3793

Qy    123 CCCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCT 182
Db   3794 CCCGTAATTCCCAACTAATGGATACGTAAAAGTAATCATTTGTCCATTAAAACTTTGGTC 3853

Qy    183 CGGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATG 242
Db   3854 AGTAATTGCTTCTTGGTAACCTGTCATGCCAGTAGTAAAAACAACTTCTCCTACTACATT 3913

Qy    243 GACAACTGTTTGGACTGATGGACTTACCAGTCTTGATC 280
Db   3914 GCCTTCTGCACCAAACGCTTTTCCTTCAAATACTGTTC 3951
```

```
RESULT 4
US-10-091-504-2110/c
; Sequence 2110, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007C1
;  CURRENT APPLICATION NUMBER: US/10/091,504
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2442
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2110
;   LENGTH: 22008
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-091-504-2110

  Query Match             8.3%;  Score 34.8;  DB 9;  Length 22008;
  Best Local Similarity   52.8%;  Pred. No. 3.7;
  Matches   75;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||  |    | ||  || |   |||| |  ||||| |   ||||  | ||| |
Db   5572 GGGGTGTCTGTGGCAGCTGCTGCTTCTGCTGCTGCTGCCGTCCCTTCCGCCGGTTCTCAA 5513

Qy    244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          | ||| | | || ||||  ||  | |  | | | ||||  |       |||| | | ||
Db   5512 ATAACAGCTGGGCCTGACTGATCTCCAGGGCCTCATCGGCATCGTCCTCGATCCCAGCAA 5453

Qy    304 ATCGAGCCCGTTCCTGGGGACC 325
          | |   || ||    || || |
Db   5452 AGCTGCCCTGTGATAGGAGAAC 5431


RESULT 5
US-09-764-869-2110/c
; Sequence 2110, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007
;  CURRENT APPLICATION NUMBER: US/09/764,869
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2442
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2110
;   LENGTH: 22008
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-869-2110

  Query Match             8.3%;  Score 34.8;  DB 10;  Length 22008;
  Best Local Similarity   52.8%;  Pred. No. 3.7;
  Matches   75;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          |||||  |    | ||  || |   |||| |  ||||| |   ||||  | ||| |
Db   5572 GGGGTGTCTGTGGCAGCTGCTGCTTCTGCTGCTGCTGCCGTCCCTTCCGCCGGTTCTCAA 5513

Qy    244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          | ||| | | || ||||  ||  | |  | | | ||||  |       |||| | | ||
Db   5512 ATAACAGCTGGGCCTGACTGATCTCCAGGGCCTCATCGGCATCGTCCTCGATCCCAGCAA 5453

Qy    304 ATCGAGCCCGTTCCTGGGGACC 325
          | |   || ||    || || |
Db   5452 AGCTGCCCTGTGATAGGAGAAC 5431


RESULT 6
US-10-081-563-25
; Sequence 25, Application US/10081563
; Publication No. US20020187495A1
; GENERAL INFORMATION:
;  APPLICANT: Wake Forest University
;  APPLICANT:  Herrington, David M.
;  APPLICANT:  Howard, Timothy D.
;  APPLICANT:  Hawkins, Gregory A.
;  APPLICANT:  Meyers, Deborah A.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS OF ESTROGEN RECEPTOR ALPHA ASSOCIATED WI
;  TITLE OF INVENTION:  FAVORABLE HDL CHOLESTEROL RESPONSE TO HORMONE REPLACEMENT THE
;  FILE REFERENCE: 9151-15
;  CURRENT APPLICATION NUMBER: US/10/081,563
;  CURRENT FILING DATE:  2002-02-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 42000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(42000)
;   OTHER INFORMATION: Nucleotides c5435870 - 53932871 of NT_023451.8
;   NAME/KEY: exon
;   LOCATION: (3372)..(3823)
;   OTHER INFORMATION: Exon 1
;   NAME/KEY: Intron
;   LOCATION: (3824)..(38055)
;   OTHER INFORMATION: Intron 1
;   NAME/KEY: exon
;   LOCATION: (38056)..(38246)
;   OTHER INFORMATION: Exon 2
US-10-081-563-25

  Query Match             8.2%;  Score 34.4;  DB 9;  Length 42000;
  Best Local Similarity   60.9%;  Pred. No. 7.2;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy     23 TCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGAT 82
          |||||||  ||| |||| ||  ||| ||  | |  ||   ||| |  | |   |||    |
Db  30973 TCATGAGATGTGGGGAAGGTGTTACGTATCTAAGAATGCAAAGTTGAACCCCTGTGAACT 31032

Qy     83 TTAAAGCTGGTGTTAAGGATTATAAATTGACT 114
          |||  ||   | ||||   |||||||| ||
Db  31033 GTAATACTTAAGATAAGTCGTATAAATTGTCT 31064


RESULT 7
US-09-933-267A-1
; Sequence 1, Application US/09933267A
; Patent No. US20020123095A1
; GENERAL INFORMATION:
;  APPLICANT: Kalush, Francis et al.
;  TITLE OF INVENTION: Estrogen receptor alpha variants and
;  TITLE OF INVENTION:  methods of detection thereof
;  FILE REFERENCE: CL000258CI4
;  CURRENT APPLICATION NUMBER: US/09/933,267A
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 60/160626
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: 60/183756
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: 09/692414
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 09/768184
;  PRIOR FILING DATE: 2001-01-24
;  PRIOR APPLICATION NUMBER: 09/804076
;  PRIOR FILING DATE: 2001-03-13
;  PRIOR APPLICATION NUMBER: 09/826314
;  PRIOR FILING DATE: 2001-04-05
;  NUMBER OF SEQ ID NOS: 2
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 465237
;   TYPE: DNA
;   ORGANISM: human
US-09-933-267A-1

  Query Match             8.2%;  Score 34.4;  DB 10;  Length 465237;
  Best Local Similarity   60.9%;  Pred. No. 28;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      23 TCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGAT 82
           ||||||| ||| |||| ||  ||| ||  | |  ||   ||| |  | |   |||    |
Db 197829 TCATGAGATGTGGGGAAGGTGTTACGTATCTAAGAATGCAAAGTTGAACCCCTGTGAACT 197888

Qy      83 TTAAAGCTGGTGTTAAGGATTATAAATTGACT 114
             |||  ||   | ||||   |||||||||  ||
Db 197889 GTAATACTTAAGATAAGTCGTATAAATTGTCT 197920


RESULT 8
US-09-952-267-6
Sequence 6, Application US/09952267
Publication No. US20030032772A1
GENERAL INFORMATION:
;  APPLICANT: HANSEN, ERIC J.
;  APPLICANT:  AEBI, CHRISTOPH
;  APPLICANT:  COPE, LESLIE D.
;  APPLICANT:  MACIVER, ISOBEL
;  APPLICANT:  FISKE, MICHAEL J.
;  APPLICANT:  FREDENBURG, ROSS A.
;  TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
;  FILE REFERENCE: AMCY:024
;  CURRENT APPLICATION NUMBER: US/09/952,267
;  CURRENT FILING DATE:  2001-09-12
;  PRIOR APPLICATION NUMBER: 09/336,447
;  PRIOR FILING DATE: 1999-06-21
;  NUMBER OF SEQ ID NOS: 98
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 3381
;   TYPE: DNA
;   ORGANISM: Moraxella catarrhalis
US-09-952-267-6

  Query Match             8.1%;  Score 34.2;  DB 9;  Length 3381;
  Best Local Similarity   49.7%;  Pred. No. 2;
  Matches   87;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy      61 GAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTAC 120
           | |  |||    |||| ||  |  | || || |||||| |  |    |  |   |  |
Db    1272 GCAGGTAATATCAGTGCTGATTCAACAGATGCTGTTAATGGCTCACAGCTATATGCTTTG 1331

Qy     121 ACCCCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAG 180
           || | |    | || || | |    ||| | |||   |||| ||     |||||   |
Db    1332 GCCACAGCTGTCAAAGCCGATGCCGATGAAAACTTTAAAGCACTCACCAAAACTCAAAAT 1391

Qy     181 CTCGGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTG 235
             | ||  ||  |  ||||||  ||  ||| ||   ||  ||  |   ||||
Db    1392 ACTTTGATTGAGCAAGGTGAAGCACAAGACGCATTAATCGCTCAAAATCAAACTG 1446


RESULT 9
US-10-092-154-1228/c
; Sequence 1228, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009C1
;  CURRENT APPLICATION NUMBER: US/10/092,154
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2003
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1228
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-092-154-1228

  Query Match             7.8%;  Score 32.8;  DB 9;  Length 495;
  Best Local Similarity   56.5%;  Pred. No. 1.9;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      44 TTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATT 103
           | || | |  | | ||| | |||||||  |    || ||  ||| |   |||  ||   ||
Db    483 TAATCTAATGATACACAAATACTAAAGTAGAAATTTGAACTAAGGGCTGTGAGAATTCTT 424

Qy     104 ATAAATTGACTTACTACACCCCGGAGTACGAAACCAAGGATACTGATA 151
           | |||  | ||| | || |  |   |||   ||   ||| | || |||
Db     423 ACAAAGAGTCTTTCAACTCTGCTCTGTAATTAAGGCAGGCTTCTCATA 376


RESULT 10
US-10-092-154-1229/c
; Sequence 1229, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009C1
;  CURRENT APPLICATION NUMBER: US/10/092,154
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2003
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1229
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-092-154-1229

  Query Match             7.8%;  Score 32.8;  DB 9;  Length 495;
  Best Local Similarity   56.5%;  Pred. No. 1.9;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      44 TTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATT 103
           | || | |  | | ||| | |||||||  |    || ||  ||| |   |||  ||   ||
Db    483 TAATCTAATGATACACAAATACTAAAGTAGAAATTTGAACTAAGGGCTGTGAGAATTCTT 424

Qy     104 ATAAATTGACTTACTACACCCCGGAGTACGAAACCAAGGATACTGATA 151
           | |||  | ||| | || |  |   |||   ||   ||| | || |||
Db     423 ACAAAGAGTCTTTCAACTCTGCTCTGTAATTAAGGCAGGCTTCTCATA 376


RESULT 11
US-09-764-847-1228/c
; Sequence 1228, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1228
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1228

  Query Match             7.8%;  Score 32.8;  DB 10;  Length 495;
```

```
  Best Local Similarity   56.5%;  Pred. No. 1.9;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      44 TTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATT 103
           | || | |  | | ||| | ||||||| |    || ||  ||| |   |||  ||   ||
Db     483 TAATCTAATGATACACAAATACTAAAGTAGAAATTTGAACTAAGGGCTGTGAGAATTCTT 424

Qy     104 ATAAATTGACTTACTACACCCCGGAGTACGAAACCAAGGATACTGATA 151
           | |||  | ||| | || |  |   |||   ||   ||| | || |||
Db     423 ACAAAGAGTCTTTCAACTCTGCTCTGTAATTAAGGCAGGCTTCTCATA 376


RESULT 12
US-09-764-847-1229/c
; Sequence 1229, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1229
;   LENGTH: 495
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1229

  Query Match             7.8%;  Score 32.8;  DB 10;  Length 495;
  Best Local Similarity   56.5%;  Pred. No. 1.9;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      44 TTATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATT 103
           | || | |  | | ||| | ||||||| |    || ||  ||| |   |||  ||   ||
Db     483 TAATCTAATGATACACAAATACTAAAGTAGAAATTTGAACTAAGGGCTGTGAGAATTCTT 424

Qy     104 ATAAATTGACTTACTACACCCCGGAGTACGAAACCAAGGATACTGATA 151
           | |||  | ||| | || |  |   |||   ||   ||| | || |||
Db     423 ACAAAGAGTCTTTCAACTCTGCTCTGTAATTAAGGCAGGCTTCTCATA 376


RESULT 13
US-09-738-626-1030
Sequence 1030, Application US/09738626
Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1030
;   LENGTH: 2373
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1030

  Query Match             7.8%;  Score 32.6;  DB 9;  Length 2373;
  Best Local Similarity   47.7%;  Pred. No. 5.4;
  Matches   95;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy      88 GCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAAACCAAGGATACT 147
           | |||||| ||| ||  | |  | ||| |   |||   | | | ||| |||  | | |
Db    1522 GATGGTGTGAAGAATGCTCAGATCACTCAGACCACGGAGAATTTCGATACCGCGCAGATC 1581

Qy     148 GATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTCGGGGTTCCGCCTGAAGAAGCAGGA 207
           ||     ||| |||    |  |    | ||| |     |  | ||   | | |
Db    1582 CTGATTACCCCAGAATTTGATGCGATCGATGAGCGCACCTCTGAGACTCTCGCAACTCTT 1641

Qy     208 GCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGGACAACTGTTTGGACTGATGGACTT 267
           |||||  |||| |   |  ||  ||  ||  |  | ||   | | | |  ||| |
Db    1642 CGTGCAGATGCTGAGACCTTCGCTGATGACACCGGCGCGACGTATGGCATTACTGGCGTC 1701

Qy     268 ACCAGTCTTGATCGTTACA 286
           |||   || |   | |||
Db    1702 ACCCCAATTTACGATGACA 1720


RESULT 14
US-09-938-842A-2369
; Sequence 2369, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2369
;   LENGTH: 2406
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-2369

  Query Match             7.6%;  Score 31.8;  DB 9;  Length 2406;
  Best Local Similarity   59.3%;  Pred. No. 9.7;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy     305 TCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACC 364
           ||| | |  || ||||| |       | | ||| || || ||| |    | ||   ||||
Db    1394 TCGTGTCTTTTTCTGGGAAGTACTTGCCACATAACTATTCTGTTGTACCTGCACCGGACC 1453

Qy     365 TATTTGAAGAGGGTTCTTGTACTAACATGTT 395
              ||   || ||||||||| ||||| |||
Db    1454 AGGATGGGAAGCGTTCTTGTAATAACAAGTT 1484


RESULT 15
US-09-924-400-251
; Sequence 251, Application US/09924400
; Patent No. US20020165371A1
; GENERAL INFORMATION:
;  APPLICANT: Frudakis, Tony N.
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Smith, John M.
```

; APPLICANT: Misher, Lynda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; APPLICANT: Li, Samuel X.
; APPLICANT: Deng, Ta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C12
; CURRENT APPLICATION NUMBER: US/09/924,400
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 251
; LENGTH: 329
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-924-400-251

uery Match 7.4%; Score 31.2; DB 9; Length 329;
est Local Similarity 54.3%; Pred. No. 4.9;
Matches 63; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy     261 TGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCACATCGAGCCCGTTCCTGG 320
           |||| |     |   | ||| || |  | |||  || || || ||  ||||||| | |
Db     206 TGGAGTGGATGGGGCTCATCTTTCCTGATGACTCTGATACCAGATACAGCCCGTCCTTCC 265

Qy     321 GGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGACCTATTTGAAGAGG 376
               |||| |||  || || ||     ||   |||       ||||| || || ||
Db     266 AAGGCCAGGTCACCATCTCAGTCGATAAGTCCATCAGCACCGCCTATCTGCAGTGG 321
```

Search completed: April 15, 2003, 18:29:19
Job time : 301 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 11:57:09 ; Search time 25043 Seconds
(without alignments)
271.617 Million cell updates/sec

Title: US-09-696-664A-3
Perfect score: 420
Sequence: 1 gcgcgttgtgagaattctta.....tcattgtgggtaacgtattt 420

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	412.2	98.1	465	12	BG159568	BG159568 OV2_6_B06
2	412.2	98.1	488	10	AW285810	AW285810 LG1_237_A
3	412.2	98.1	490	10	BE598732	BE598732 PI1_81_A0
4	412.2	98.1	490	12	BG053104	BG053104 RHIZ2_16_
5	412.2	98.1	492	12	BG559375	BG559375 RHIZ2_54_
6	412.2	98.1	495	10	AW563595	AW563595 LG1_241_C
7	412.2	98.1	496	12	BG102147	BG102147 RHIZ2_21_
8	412.2	98.1	517	13	BM323998	BM323998 PIC1_30_B
9	412.2	98.1	536	12	BG465717	BG465717 RHIZ2_47_
10	412.2	98.1	546	10	BE596352	BE596352 PI1_49_G0
11	412.2	98.1	551	10	BE596365	BE596365 PI1_52_A0
12	412.2	98.1	552	12	BF705205	BF705205 RHIZ2_2_A
13	412.2	98.1	565	13	BM323693	BM323693 PIC1_22_C
14	412.2	98.1	576	10	AW671847	AW671847 LG1_352_C
15	412.2	98.1	576	12	BG102878	BG102878 RHIZ2_34_
16	412.2	98.1	582	12	BF657539	BF657539 OV2_23_E0
17	412.2	98.1	593	13	BM326083	BM326083 PIC1_65_E
18	412.2	98.1	598	13	BG948809	BG948809 LG1_266_E
19	412.2	98.1	599	13	BM326559	BM326559 PIC1_58_E
20	412.2	98.1	601	13	BM326409	BM326409 PIC1_56_F
21	412.2	98.1	614	10	BE594480	BE594480 PI1_33_H0
22	412.2	98.1	615	10	BE599555	BE599555 PI1_88_H0
23	412.2	98.1	630	12	BF704811	BF704811 FM1_51_G1
24	412.2	98.1	630	13	BM325003	BM325003 PIC1_38_E
25	411	97.9	515	13	BM325550	BM325550 PIC1_46_A
26	411	97.9	627	13	BM325185	BM325185 PIC1_41_D
27	410.6	97.8	583	10	BE592098	BE592098 LG1_224_C
28	410.4	97.7	596	12	BG357871	BG357871 OV2_31_C0
29	408.2	97.2	444	12	BG560424	BG560424 RHIZ2_74_
30	408.2	97.2	450	12	BG241641	BG241641 RHIZ2_49_
31	408.2	97.2	487	12	BG465536	BG465536 RHIZ2_45_
32	408.2	97.2	487	13	BM323586	BM323586 PIC1_21_A
33	408.2	97.2	488	10	AW671318	AW671318 LG1_335_B
34	408.2	97.2	490	12	BG159586	BG159586 OV2_6_D09
35	408.2	97.2	491	10	AW564022	AW564022 LG1_281_G
36	408.2	97.2	491	12	BG465558	BG465558 RHIZ2_46_
37	408.2	97.2	493	12	BG559241	BG559241 RHIZ2_52_
38	408.2	97.2	513	13	BM322939	BM322939 PIC1_15_D
39	408.2	97.2	514	13	BM323531	BM323531 PIC1_20_D
40	408.2	97.2	562	12	BG052307	BG052307 RHIZ2_12_
41	408.2	97.2	562	13	BM324614	BM324614 PIC1_33_D
42	408.2	97.2	572	12	BG488245	BG488245 RHIZ2_61_
43	406.6	96.8	487	10	AW671457	AW671457 LG1_347_B
44	406.6	96.8	487	10	BE597559	BE597559 PI1_71_B0
45	406.6	96.8	488	10	AW286737	AW286737 LG1_204_G

ALIGNMENTS

RESULT 1
BG159568
LOCUS BG159568 465 bp mRNA linear EST 06-FEB-2001
DEFINITION OV2_6_B06.b1_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG159568
VERSION BG159568.1 GI:12693232
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 465)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 435
POLYA=No.
FEATURES Location/Qualifiers

```
    source          1. .465
                    /organism="Sorghum bicolor"
                    /db_xref="taxon:4558"
                    /clone_lib="Ovary 2 (OV2)"
                    /note="Organ: Mix of ovaries of varying immature stages
                    from 8-week-old plants; Vector: pBluescript II from Lambda
                    Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                    from poly-A RNA in the cloning vector lambda ZAP II.
                    Clones to be sequenced were prepared by mass excision."
BASE COUNT      124 a     94 c    113 g    134 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 12;  Length 465;
  Best Local Similarity   99.3%;  Pred. No. 4.7e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 2
AW285810
LOCUS       AW285810                488 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  LG1_237_A11.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW285810
VERSION     AW285810.2  GI:6858276
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6675654.
            Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 487
            POLYA=No.
FEATURES            Location/Qualifiers
     source         1. .488
                    /organism="Sorghum bicolor"
                    /db_xref="taxon:4558"
                    /clone_lib="Light Grown 1 (LG1)"
                    /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                    seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                    ; The library was made from poly-A RNA in the cloning
                    vector lambda ZAP II.  Clones to be sequenced were
                    prepared by mass excision."
BASE COUNT      131 a    101 c    112 g    144 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 10;  Length 488;
  Best Local Similarity   99.3%;  Pred. No. 4.8e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 62

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 122

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 182

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 242

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 302

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 362

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     363 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 419


RESULT 3
BE598732
LOCUS       BE598732                490 bp    mRNA    linear   EST 18-AUG-2000
DEFINITION  PI1_81_A06.b1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE598732
VERSION     BE598732.1  GI:9853805
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
```

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 465
POLYA=No.

FEATURES Location/Qualifiers
source 1. .490
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (PI1)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate FRM42I of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptiblity to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 132 a 100 c 116 g 141 t 1 others
ORIGIN

Query Match 98.1%; Score 412.2; DB 10; Length 490;
Best Local Similarity 99.3%; Pred. No. 4.8e-118;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427
```

RESULT 4
BG053104
LOCUS BG053104 490 bp mRNA linear EST 25-JAN-2001
DEFINITION RHIZ2_16_F07.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA sequence.
ACCESSION BG053104
VERSION BG053104.1 GI:12508448
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 490)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 463
POLYA=No.

FEATURES Location/Qualifiers
source 1. .490
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 132 a 101 c 116 g 141 t
ORIGIN

Query Match 98.1%; Score 412.2; DB 12; Length 490;
Best Local Similarity 99.3%; Pred. No. 4.8e-118;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427
```

RESULT 5
BG559375
LOCUS BG559375 492 bp mRNA linear EST 10-APR-2001
DEFINITION RHIZ2_54_A11.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA sequence.
ACCESSION BG559375
VERSION BG559375.1 GI:13588373
KEYWORDS EST.

```
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
            High quality sequence stop: 487
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      133 a    102 c    116 g    141 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 12;  Length 492;
  Best Local Similarity   99.3%;  Pred. No. 4.8e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 6
AW563595
LOCUS       AW563595                 495 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  LG1_241_C12.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW563595
VERSION     AW563595.1  GI:7217473
KEYWORDS    EST.
```

```
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 476
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      133 a    102 c    116 g    144 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 10;  Length 495;
  Best Local Similarity   99.3%;  Pred. No. 4.8e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 7
BG102147
LOCUS       BG102147                 496 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  RHIZ2_21_D07.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
```

```
ACCESSION   BG102147
VERSION     BG102147.1  GI:12616980
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
            High quality sequence stop: 488
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      133 a    103 c    116 g    144 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 12;  Length 496;
  Best Local Similarity   99.3%;  Pred. No. 4.8e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 8
BM323998
LOCUS       BM323998                517 bp    mRNA    linear   EST 04-JAN-2002
DEFINITION  PIC1_30_B01.b1_A002 Pathogen-infected compatible 1 (PIC1) Sorghum
            bicolor cDNA, mRNA sequence.
ACCESSION   BM323998
VERSION     BM323998.1  GI:18062304
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R.,
            Sudman,M. and Pratt,L.H.
  TITLE     An EST database from Sorghum: plants infected with a compatible
            pathogen
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector, and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20. Three-prime sequences, which are obtained with PolyTMix or
            T7 sequencing primer, are presented as the reverse complement.
            Seq primer: JEN REV
            High quality sequence stop: 491
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Sorghum bicolor"
                     /cultivar="BTx623"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen-infected compatible 1 (PIC1)"
                     /tissue_type="Leaves"
                     /dev_stage="4-week-old seedlings infected with
                     Colletotrichum graminicola"
                     /note="Vector: pBluescript II SK(-) from Lambda Zap II;
                     Site_1: XhoI; Site_2: EcoRI; Four-week-old sorghum
                     seedlings were sprayed with spore suspension prepared from
                     3-week-old FRM42I, a sorghum isolate of the anthracnose
                     pathogen Colletotrichum graminicola. Inoculated plants
                     were kept in a 25 C dark growth chamber with 100% relative
                     humidity for 24 hr, followed by 12/12 hr of light/dark
                     cycle at 25 C with 90% relative humidity for another 24
                     hr. All leaves were harvested and quick frozen with liquid
                     nitrogen and stored in a -80 C freezer. The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision.
                     WARNING: While most or all ESTs are expected to derive
                     from the host plant, no effort was made to eliminate ESTs
                     deriving from the pathogen."
BASE COUNT      137 a    111 c    121 g    148 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 13;  Length 517;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy    244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy    304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy    364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db    371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 9
BG465717
LOCUS       BG465717                536 bp    mRNA    linear   EST 20-MAR-2001
DEFINITION  RHIZ2_47_G09.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG465717
VERSION     BG465717.1  GI:13394693
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
            High quality sequence stop: 526
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      145 a    112 c    127 g    152 t
ORIGIN

  Query Match           98.1%;  Score 412.2;  DB 12;  Length 536;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy     64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy    124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy    184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy    244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy    304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy    364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db    371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 10
BE596352
LOCUS       BE596352                546 bp    mRNA    linear   EST 18-AUG-2000
DEFINITION  PI1_49_G09.b1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE596352
VERSION     BE596352.1  GI:9851425
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 494
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen induced 1 (PI1)"
                     /note="Organ: Anthracnose-infected leaves from
                     two-week-old sorghum plants 48 hr after inoculation;
                     Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                     cultivar) were infected with pathogen (isolate FRM42I of
                     Colletotrichum graminicola, which is a sorghum isolate).
                     RNA was prepared from infected leaves harvested from 45
                     seedlings  48 hours after inoculation.  Note: young
                     seedlings (2 weeks old) exhibit juvenile resistant
                     reaction, which is an incompatible interaction. As they
                     grow older (4 weeks or older), plants resume susceptiblity
                     to anthracnose disease. The library was made from poly-A
                     RNA in the cloning vector lambda ZAP II. Clones to be
                     sequenced were prepared by mass excision. WARNING: While
                     most or all ESTs are expected to derive from the host
                     plant, no effort was made to eliminate ESTs deriving from
                     the pathogen."
BASE COUNT      151 a    113 c    128 g    154 t
ORIGIN

  Query Match           98.1%;  Score 412.2;  DB 10;  Length 546;
  Best Local Similarity   99.3%;  Pred. No. 5e-118;
```

```
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 11
BE596365
LOCUS       BE596365                551 bp    mRNA    linear   EST 18-AUG-2000
DEFINITION  PI1_52_A04.b1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE596365
VERSION     BE596365.1  GI:9851438
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 495
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen induced 1 (PI1)"
                     /note="Organ: Anthracnose-infected leaves from
                     two-week-old sorghum plants 48 hr after inoculation;
                     Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                     cultivar) were infected with pathogen (isolate FRM42I of
                     Colletotrichum graminicola, which is a sorghum isolate).
                     RNA was prepared from infected leaves harvested from 45
                     seedlings  48 hours after inoculation.  Note: young
                     seedlings (2 weeks old) exhibit juvenile resistant
                     reaction, which is an incompatible interaction. As they
                     grow older (4 weeks or older), plants resume susceptiblity
                     to anthracnose disease. The library was made from poly-A
                     RNA in the cloning vector lambda ZAP II. Clones to be
                     sequenced were prepared by mass excision. WARNING: While
                     most or all ESTs are expected to derive from the host
                     plant, no effort was made to eliminate ESTs deriving from
                     the pathogen."
BASE COUNT      150 a    115 c    131 g    155 t
ORIGIN

  Query Match           98.1%;  Score 412.2;  DB 10;  Length 551;
  Best Local Similarity   99.3%;  Pred. No. 5e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 12
BF705205
LOCUS       BF705205                552 bp    mRNA    linear   EST 29-DEC-2000
DEFINITION  RHIZ2_2_A03.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BF705205
VERSION     BF705205.1  GI:11996866
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
```

```
            High quality sequence stop: 502
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .552
                      /organism="Sorghum propinquum"
                      /db_xref="taxon:132711"
                      /clone_lib="Rhizome2 (RHIZ2)"
                      /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                      Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                      from poly-A RNA in the cloning vector lambda ZAP II.
                      Clones to be sequenced were prepared by mass excision."
BASE COUNT      151 a    115 c    131 g    155 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 12;  Length 552;
  Best Local Similarity   99.3%;  Pred. No. 5e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 13
BM323693
LOCUS       BM323693                565 bp    mRNA    linear   EST 04-JAN-2002
DEFINITION  PIC1_22_C01.b1_A002 Pathogen-infected compatible 1 (PIC1) Sorghum
            bicolor cDNA, mRNA sequence.
ACCESSION   BM323693
VERSION     BM323693.1  GI:18061719
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R.,
            Sudman,M. and Pratt,L.H.
  TITLE     An EST database from Sorghum: plants infected with a compatible
            pathogen
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu

            Sequences have been trimmed to exclude PolyA, vector, and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20. Three-prime sequences, which are obtained with PolyTMix or
            T7 sequencing primer, are presented as the reverse complement.
            Seq primer: JEN REV
            High quality sequence stop: 500
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .565
                      /organism="Sorghum bicolor"
                      /cultivar="BTx623"
                      /db_xref="taxon:4558"
                      /clone_lib="Pathogen-infected compatible 1 (PIC1)"
                      /tissue_type="Leaves"
                      /dev_stage="4-week-old seedlings infected with
                      Colletotrichum graminicola"
                      /note="Vector: pBluescript II SK(-) from Lambda Zap II;
                      Site_1: XhoI; Site_2: EcoRI; Four-week-old sorghum
                      seedlings were sprayed with spore suspension prepared from
                      3-week-old FRM42I, a sorghum isolate of the anthracnose
                      pathogen Colletotrichum graminicola. Inoculated plants
                      were kept in a 25 C dark growth chamber with 100% relative
                      humidity for 24 hr, followed by 12/12 hr of light/dark
                      cycle at 25 C with 90% relative humidity for another 24
                      hr. All leaves were harvested and quick frozen with liquid
                      nitrogen and stored in a -80 C freezer. The library was
                      made from poly-A RNA in the cloning vector lambda ZAP II.
                      Clones to be sequenced were prepared by mass excision.
                      WARNING: While most or all ESTs are expected to derive
                      from the host plant, no effort was made to eliminate ESTs
                      deriving from the pathogen."
BASE COUNT      152 a    117 c    135 g    161 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 13;  Length 565;
  Best Local Similarity   99.3%;  Pred. No. 5e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
           |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


RESULT 14
AW671847
LOCUS       AW671847                576 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  LG1_352_C10.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW671847
```

```
VERSION     AW671847.1  GI:7535747
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 490
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      156 a    112 c    122 g    186 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 10;  Length 576;
  Best Local Similarity   99.3%;  Pred. No. 5.1e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 190

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 250

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 310

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 370

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 430

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     431 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 490

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     491 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 547


RESULT 15
BG102878
LOCUS       BG102878                 576 bp    mRNA    linear   EST 30-JAN-2001
```

```
DEFINITION  RHIZ2_34_E06.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG102878
VERSION     BG102878.1  GI:12617711
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
            High quality sequence stop: 494
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      156 a    118 c    136 g    166 t
ORIGIN

  Query Match             98.1%;  Score 412.2;  DB 12;  Length 576;
  Best Local Similarity   99.3%;  Pred. No. 5.1e-118;
  Matches  414;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGACTTATGTCACCACAAACAGAA 70

Qy      64 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 ACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACC 130

Qy     124 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 CCGGAGTACGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCTC 190

Qy     184 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGGGTTCCGCCTGAAGAAGCAGGAGCTGCAGTAGCTGCGGAATCTTCTACTGGTACATGG 250

Qy     244 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 ACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGACGATGCTATCAC 310

Qy     304 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 ATCGAGCCCGTTCCTGGGGACCCAGATCAATATATCTGTTATGTAGCTTATCCATTAGAC 370

Qy     364 CTATTTGAAGAGGGTTCTTGTACTAACATGTTTACTTTCATTGTGGGTAACGTATTT 420
          |||||||||||||||||  |||||||||||||||| ||||||||||||||||||||
Db     371 CTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT 427


Search completed: April 16, 2003, 01:25:10
Job time : 25049 secs
```